Event Level RICH algorithm: An explanation and user guide

HERMES Internal Note 07-017

Rebecca Lamb
rlamb2@uiuc.edu
University of Illinois at Urbana-Champaign

Achim Hillenbrand
Achim.Hillenbrand@desy.de
DESY Zeuthen

July 9, 2008

Abstract

The EVT event level RICH algorithm is explained and details of the output are explained for analyzers. The results of systematic studies done on MC are also presented. The details of the use of new Pmatrices are explained.

Contents

1 Analyzer's Summary

The information you are seeking, in its most up to date form, can probably be found on the RICH wiki page (`http://hermes-wiki.desy.de/index.php/RICH_PID`). If you want more details, read on.

1.1 Use

- **Linking to the smRICH table**

  To use EVT simply link to the EVT row of the smRICH table via the EVT link. Ultimately, this is all there is to do to use the EVT method. However, the very first productions which included the EVT method require some extra effort. For a single track in one detector half, EVT is identical to the DRT method. In the first EVT productions, the DRT method has to be selected manually for these events. In all new productions, the DRT method is linked as EVT for single tracks. See section 3.6 for a listing of the current productions with the status of the EVT single track link. See sections 3.5 for some pseudo code for use in a hanna code. See the Hanna++ RICH example for use in Hanna++. For more details on the information in the uDST files see section 3.3

- **Track counting**

  The number of tracks in a detector half includes all long tracks. Thus, when using the Pmatrices one should count the number of long tracks in the detector half before making any cuts on the tracks. For example, in constructing the Pmatrices, one may have two long tracks in a particular event: track A within all cuts and track B outside of the fiducial volume cuts. In this case track A will contribute to the **2-track** Pmatrix while track B would not contribute at all.

- **Momentum cuts**

  As can be seen in Figure 1, RICH hadron identification is only possible in a limited momentum range. Identification at very low momenta is dangerous because it depends on a lack of detection of kaon and/or proton rings, which are easily lost due to detector acceptance (low momentum tracks are strongly bent in the magnet and are pushed to the detector edge) and PMT inefficiencies (near the Čerenkov threshold very few photons are produced). The

Pmatrices provided cover the momentum range 1-16GeV, however the results are only reasonable in the range 2-15GeV. Pion only identification is possible down to 1GeV. Kaons and especially protons have higher systematic errors in the 2-4GeV range and thus, for analyses limited by the RICH systematic error, a minimum momentum cut on protons and possibly kaons of 4GeV should be investigated. Anytime analyzers choose to use only some subset of the available hadrons they should take care to use an appropriate Pmatrix. For example, for pion only identification for 1-2GeV, the standard Pmatrix provided should be converted to a 2x2 matrix with pions and non-pions (kaons and protons) and then inverted. An analogous procedure should be used for the 2-4GeV bins if the kaon/proton minimum momentum cut is changed to 4GeV. These can be constructed from the "pmatyields" files that are provided by adding together the appropriate rows and columns and then normalizing by the total in each column (including the unidentified category).

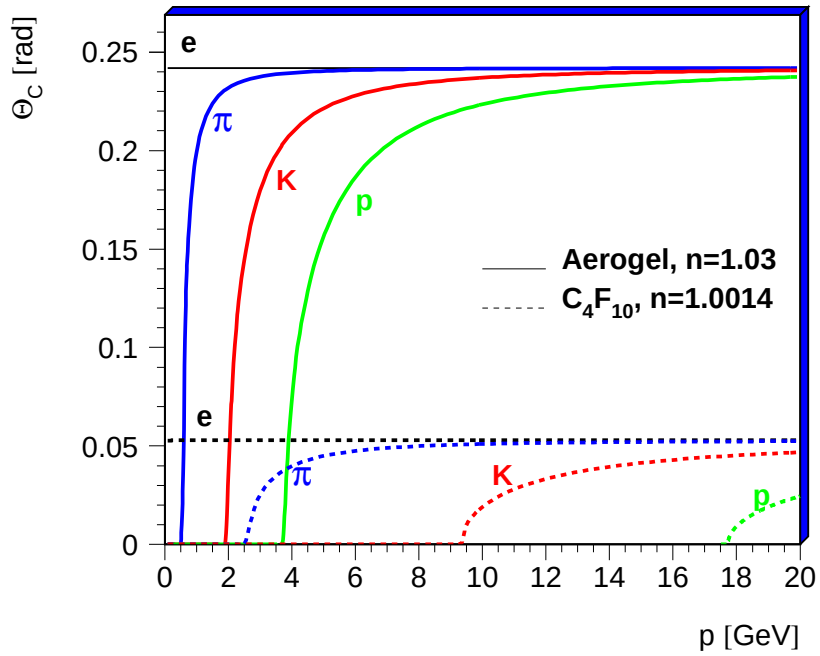

Figure 1: The Čerenkov photon opening angles for pions, kaons, and protons in the two RICH radiators, as a function of momentum.

1.2 Reasons for choosing EVT+DRT

The recommended method is DRT for single tracks and EVT for multiple tracks.

For single tracks the choice to use DRT instead of IRT was not obvious:

- **Pmatrices**

  IRT in most cases has a lower particle identification efficiency than EVT (see Figure 34). This results in larger weights for the unfolding procedure, in turn leading to an increase of the statistical error [1] .

- **Pepsi Challenge**

  IRT does better in the PEPSI challenge when using different generators for hadron yields and Pmatrices used to unfold the hadron yields. Figure 15 shows the ratio of unfolded to true yields from a disNG MC production, unfolded with Pmatrices from Pythia. However, this is the only aspect of IRT which is better than EVT. Using disNG both for yields and Pmatrices results in a better PEPSI challenge result for EVT (Figure 14).

- **10 GeV threshold effect:**

  Unlike the EVT method, the IRT method has always had problems distinguishing between kaons and protons with a momentum around 10 GeV (see IRT P matrices in Figure 34). This is the range where a signal from the C_4F_{10} gas can be expected for kaons, though the photon yield should still be low (see Figure 1). Looking at the raw (non-unfolded) K^+ multiplicities versus momentum (Figure 2), the low efficiency for IRT is clearly visible in the data. However, the size of the efficiency drop seems to be overestimated in the MC simulation, since the unfolded multiplicities (Figure 3) actually show an excess of kaons in that region for the IRT method. Conversely there is no indication of a problem for the EVT method. While the problem is only obvious for a single momentum bin, it could indicate a more complex issue with IRT.

For multiple tracks events

[1] Large off-diagonal Pmatrix elements result in increased shuffling of counts, which increases statistical error. As an (overly simplified) example consider having two tracks, a pion and a kaon. If the Pmatrix is the identity matrix (and therefore the the inverted Pmatrix is also the identity matrix), then unfolding will give 1 and 0 counts, resulting in a statistical error on the pions of $\sqrt{1^2 + 0^2} = 1$. However, a non-diagonal Pmatrix might unfold to give a pion error of $\sqrt{(1.2)^2 + (-0.2)^2} = 1.48$.

- **Pmatrices**

  EVT is better than IRT. See Figure 34 (right hand side).

- **Pepsi Challenge**

  EVT is better than IRT. See Figures 16 and 17.

- **Unphysical anti-proton yields in the right side of the RICH detector**

  As shown in Figure 4, unfolding with IRT Pmatrices results in unphysical (negative) yields of anti-protons traversing the RICH at $x < 0$. This indicates that IRT has a dependence on this variable (and potentially others), while the Pmatrices do not depend on these variables and therefor can only correct things in an averaged way. EVT does not appear to have such a strong dependence on x.

- Δx_{RICH}

  IRT has problems when two tracks are close together and their Čerenkov rings overlap, as explained in section 2.1. In Figure 5 the difference in x-distance at the RICH between the DIS lepton and a hadron is plotted when the hadron is alone in the detector half and when it is accompanied by the lepton. EVT was designed to remedy this problem. You can see in Figure 6 that the Δx_{RICH} distribution are much smoother for EVT.

1.3 Other Studies: Alignment, Lambda Pmatrices and Background files

- **Alignment**

  The detector alignment was not studied in detail. However, visual inspection of events with the Hermes Rich Event display (HeRE), showed no obvious problems. Also, as the mirror alignment was tuned to data, alignment is not thought to be a concern.

- **Lambda P matrices**

  Decaying particle Pmatrices using Lambdas for IRT (Figure 47) and EVT (Figure 48) were constructed and are described in section 6.7. Here the results seem to be more consistent between MC and data for EVT, offering another reason to support the use of EVT.

DRT and EVT now both include a PMTs dependent number that is used as the expectation of a hit in a PMT due to physics background, which are contained in the "background file".

- **Backgrounds used in productions**

  In both data productions and MC productions it is best to use a background file extracted from the production itself. This is done for all the new data productions.

- **Hot PMTs and Dead PMTs**

  Hot and dead PMTs are evaluated by looking at the total number of hits in the tube over a large data sample. They are denoted in the background file by a negative number and are skipped over when computing the likelihood. Studies of the effect of hot and dead tubes can be found in section 5.2.

- **Cuts on background files**

  Variations in the method of production the background file produce no significant effect in data sets than had stable behavior (some data sets had strange behavior, see section 4.1.1 for a summary). The final background files were produced by requiring the "background" detector half to have no tracks of any kind and the other detector half to contains a DIS lepton ($Q^2 > 10, W^2 > 10, y < 0.92$) within fiducial cuts.

- **Beampipe simulation on/off**

  A study was done to determine if the background files from MC depend on the beampipe simulation being on. No effect was found. See Figure 26 and section 5.1.1

1.4 Pmatrices and Systematic error

- **Files**

  Final Pmatrices are at `/group01/richgrp/Pmatrix_v4.0/`. The format is described in the file headers as well as on the wiki. The various files provided are described below.

- **Cuts**

  Cuts used to generate the Pmatrices can be found in table 3 page 45, with more information in section 6.2

- **Charge Separation**

  Charge separated, and "charge-likeness" separated Pmatrices (see section 6.3 for an explanation of likeness separation) were investigated, but as charge separated matrices would require

MC productions with both beam charges, and likeness separated Pmatrices make no difference for EVT (see Figure 13), it was decided to keep the method as simple as possible and use charge combined Pmatrices.

- **Systematics: backgrounds and generators**

  Since background files have been found to differ between data and MC, and between MC generators, and the choice of file has some effect on the Pmatrices it was decided to use Pmatrices with different files as a systematic error estimation. The "best guess", that is, the Pmatrix where the background file is taken from the production itself, should be used as the central value. Additionally, as the Pmatrices depend on the MC generator used, Pmatrices are provided with both disNG and pythia. The Pmatrices that are provided to the analyzer to produce a systematic error on the RICH unfolding are summarized in Table 1 and are plotted in Figures 7 to 9 for the case of the 99 geometry (data taking years 1999 to 2005) and in Figures 10 to 12 for the case of the 06 geometry (data taking years 2006/07). The Pmatrix sets show similar variations for the 99 and 06 geometries, however, in the case of 06 the variation is larger. This is consistent with the observed larger changes of the background files used for the 06 geometry (compare eg Figures 27 and 28). From Figures 7 to 12 it is also apparent that the background dependence is the major influence on the systematics, since using Pythia or disNG as the generator yields very similar results when in both cases the Pythia background is used.

Table 1: Pmatrices provided for systematic error estimation

Generator	**Background**
disNG	disNG → central value
	pythia
	data
pythia	pythia

- **e- h- and c- tunes** As the variations described above cover a larger range than the previously used e- h- and c- tune Pmatrices (see Figures 30 to 32), these have been abandoned.

- **Detector geometry**

  As it was found that the background files and Pmatrices taken from MC with 1999 geometry and from 2006 geometry (with the shifted target cell) differ significantly (see Figures 27 and

28), a full set of Pmatrices are provided separately for pre-Recoil (1999 geometry) and Recoil (2006 geometry) data. The background files used to produce these Pmatrices are taken from productions with the same geometry. Analyzers should always use the Pmatrices appropriate for the geometry of the data set they are using.

- **IRT Pmatrices**

  Pmatrices are also provided for IRT. As IRT does not depend on the background file, only two versions are provided, disNG and Pythia. However, it should be noted that the difference between these two version is an underestimate of the systematic error due to all of the problems with IRT highlighted above, particularly the 10GeV threshold "bump" in the kaon multiplicities, which is not covered by variation in the IRT pmatrices (see Figure 3).

- **old v3.0 IRT Pmatrices**

  The new IRT Pmatrices differ from the old v3.0 Pmatrices. Most importantly, in the momentum range between 2 and 4 GeV the v3.0 Pmatrices underestimate the misidentification of pions as kaons and in turn overestimate the misidentification of pions as protons when compared to the new IRT matrices (see Figure 33). Unfolding with these matrices thus leads to higher kaons yields and lower (anti-)proton yields in these two momentum bins, as can be seen in Figure 69. Note that the new IRT matrices are closer to EVT than the v3.0 matrices.

- **old systematic errors based on RICH MC tunes**

  With the old v3.0 IRT matrices, the systematic error was estimated using two alternative matrix sets ("e-tune" and "h-tune"), which represented different tunes of the RICH MC (to either electron/positron data or decaying hadron data), while the standard "center" matrices were extracted by using the middle value of the tune parameter. The new EVT P matrices were all extracted using the "e-tune" setting. While the variation of the old v3.0 IRT Pmatrices with the tune parameter (Figures 30 to 32) is at high momentum often larger than the variation of the different EVT Pmatrix sets (Figures 7 to 9), the EVT method does not appear to be so sensitive to the various tunes. This can be seen in Figure 18, showing the difference of EVT matrices extracted using the "center", "e-tune" and "h-tune" MC settings.

- **A Caution**

  Analyses that depend on variables that the Pmatrices are integrated over (such as ϕ) should do a PEPSI challenge to determine the systematic error as the systematic generated by the

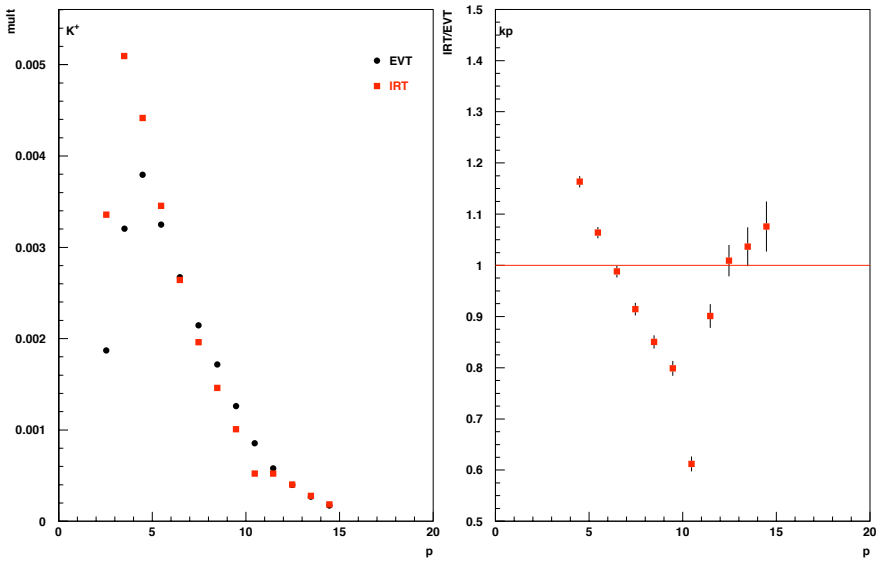

Figure 2: Raw K^+ multiplicities, as identified by the EVT and IRT method in the 00d2 data production versus momentum. The low kaon identification efficiency for IRT is clearly visible in the 10-11 GeV bin.

various generators and backgrounds may not accurately reflect the variations seen versus particular variables.

More information on the studies done concerning variation in the Pmatrices can be found in Chapter 6.

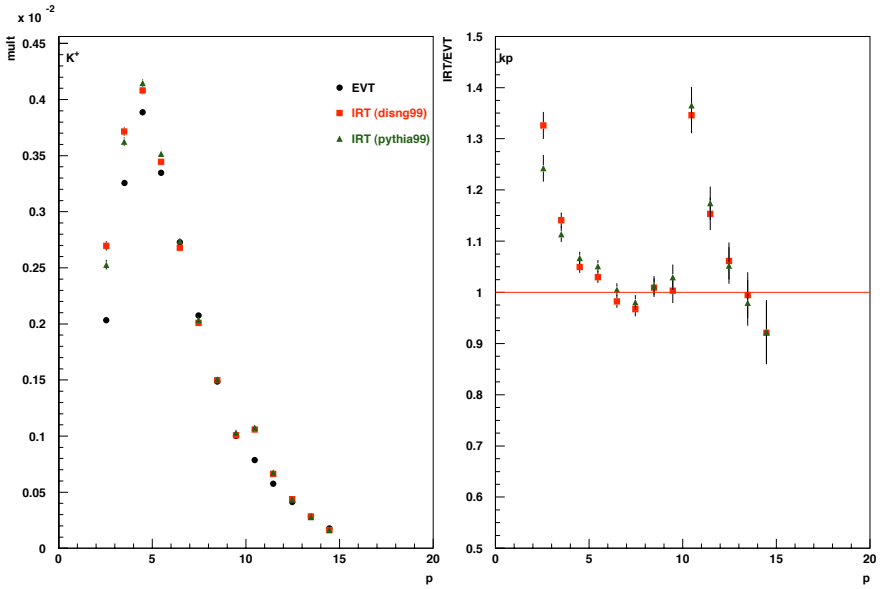

Figure 3: Unfolded K^+ multiplicities versus momentum. Unfolding the multiplicities with IRT Pmatrices (from disNG or Pythia) results in an overcorrection of the low identification efficiency.

2 Background Information: DRT

The Direct Ray Tracing (DRT) algorithm is used to determine the hadron type of tracks passing through the RICH. It is described in [1]. Basically, it performs a Monte Carlo simulation of the RICH's response to a track with the kinematics of the track in question combined with a single particle type hypothesis (PTH). Many Čerenkov photons are generated and then the hit pattern is normalized to the expected number of PMT hits for the track. A constant background number is added to the expected number of counts in each PMT. Finally, the simulated PMT hit pattern is compared to the observed hit pattern and the likelihood of such a pattern is computed. This is repeated for each PTH (pion, kaon, and proton). The output to the μDST smRICH table consists of iType, rQp, and rProb[5]. The iType is assigned to the most likely PTH, and the ratio of the first and second most likely PTH gives rQp. rProb[5] holds the computed log-likelihoods for each PTH.

Please note: HERMES lepton / hadron PID is never used in any RICH algorithm (IRT, DRT or EVT). Whether or not the electron (positron) hypothesis is considered for a given track is an

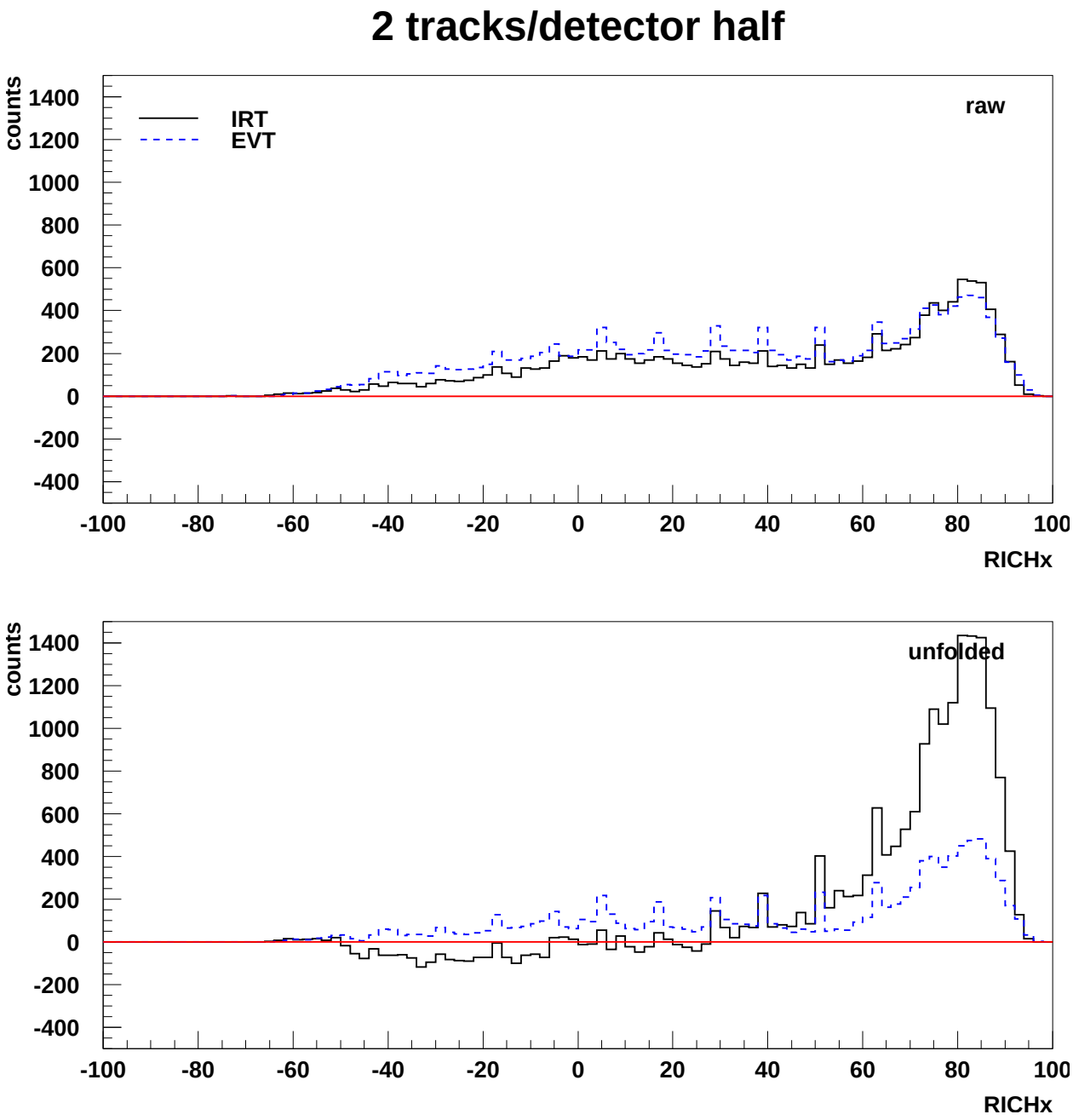

Figure 4: x distribution of antiprotons at the RICH for the case of two tracks in the detector half. Top: IRT and EVT results before unfolding. Bottom: IRT and EVT results after unfolding. EVT shows little change after unfolding. IRT has unphysical negative values after unfolding.

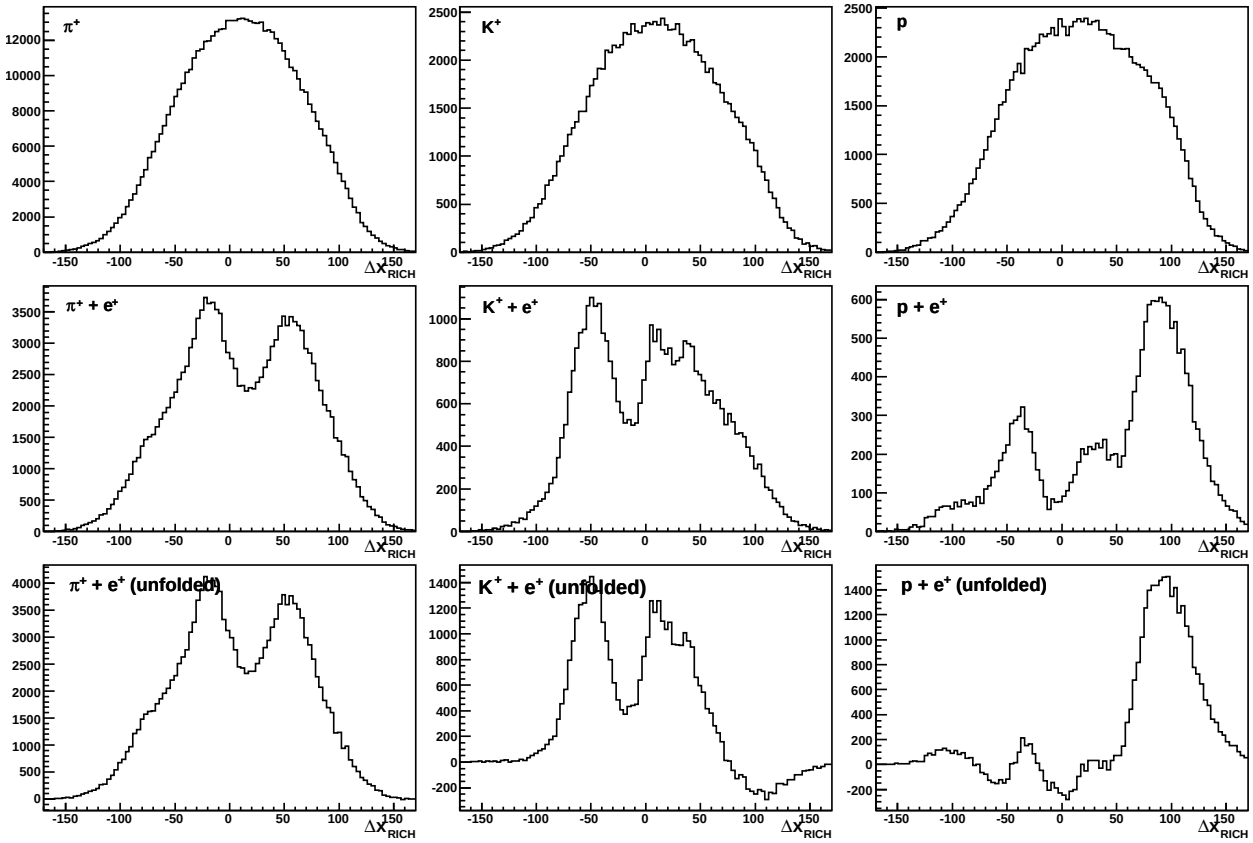

Figure 5: Δx_{RICH} is the x-distance at the RICH between the hadron and the DIS lepton. Shown are positive pions, kaons and protons (in columns) identified with IRT. The first row is the raw (not unfolded) hadron counts where the hadron is alone in the detector half. The second row is the raw count when the hadron and the lepton are in the same detector half. The bottom row is the same as the second row but after RICH unfolding.

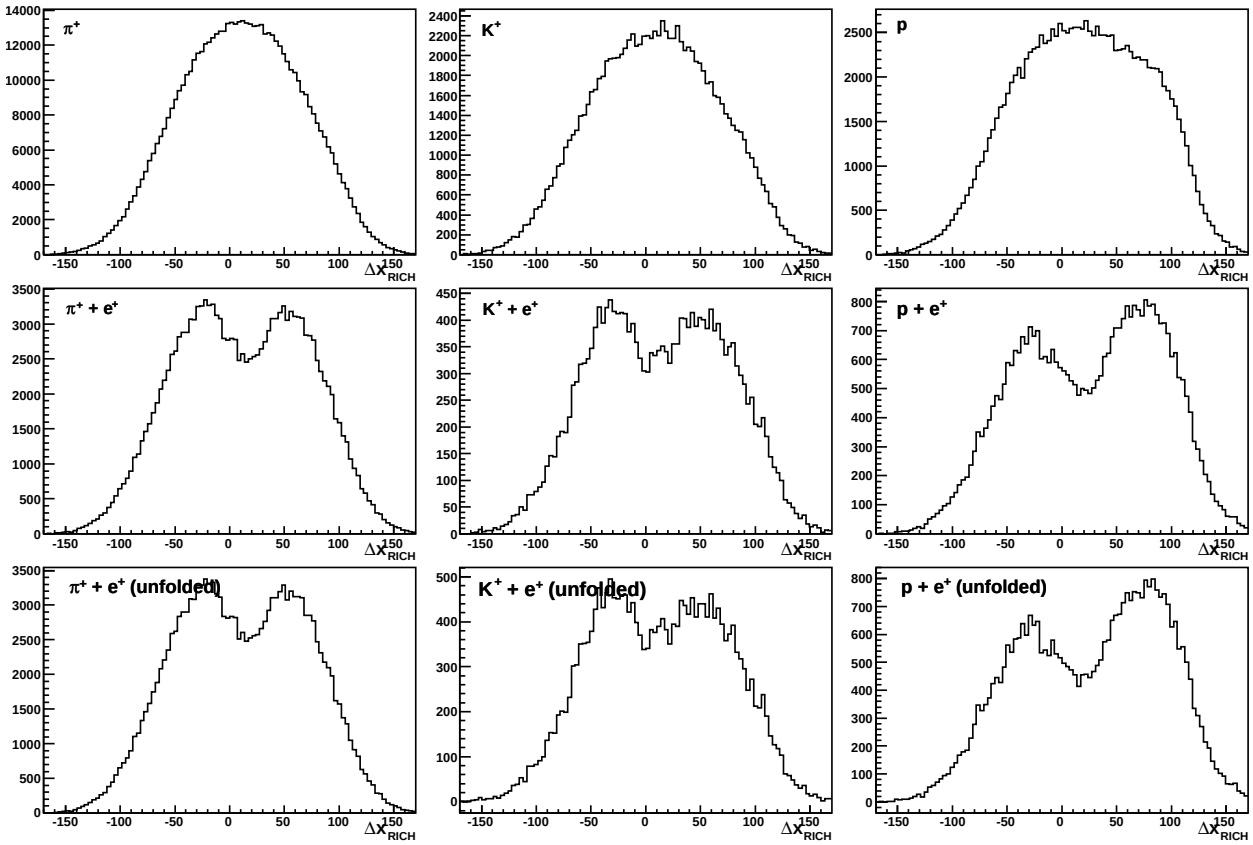

Figure 6: Same as Figure 5, but for EVT.

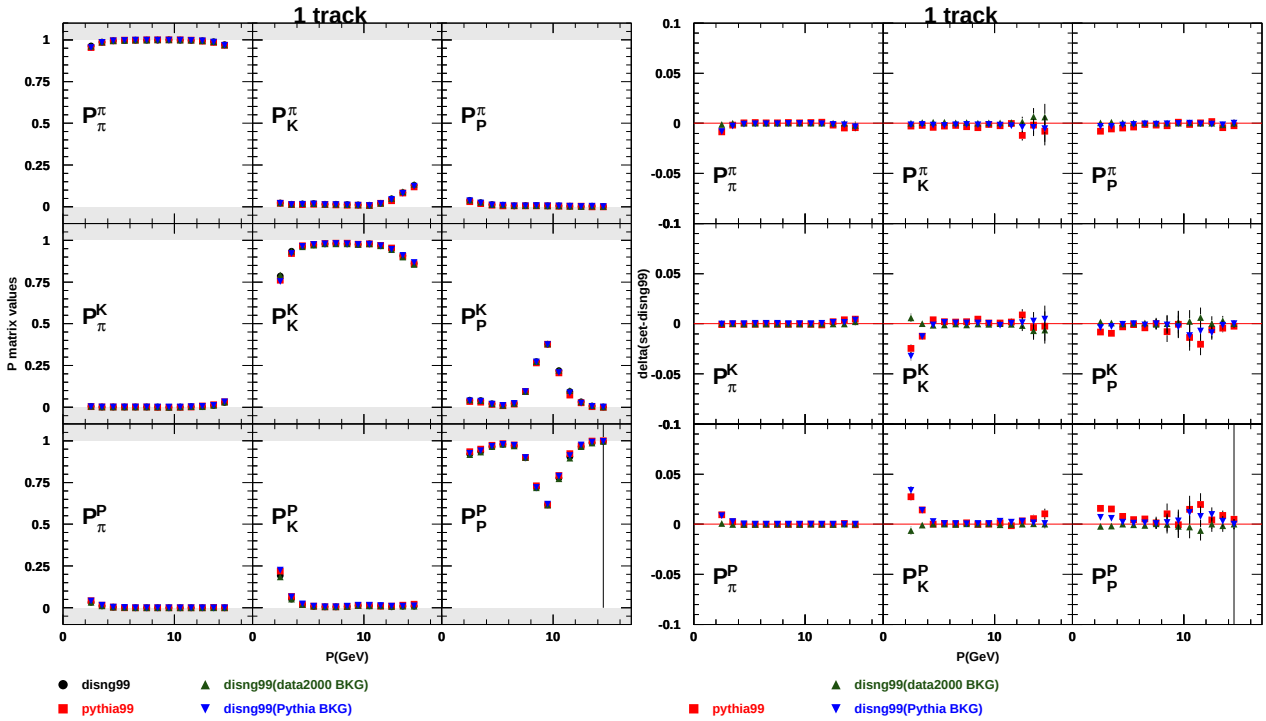

Figure 7: The four EVT Pmatrix sets for the 99 geometry and 1 track per detector half: disng99 is used for the central values, the systematic error is estimated using a different generator (pythia99) and disng99 with different background assumptions (disng99 with BKG from 2000 data and Pythia BKG, respectively). The plot on the right shows the difference of the Pmatrix values to disng99 (own BKG).

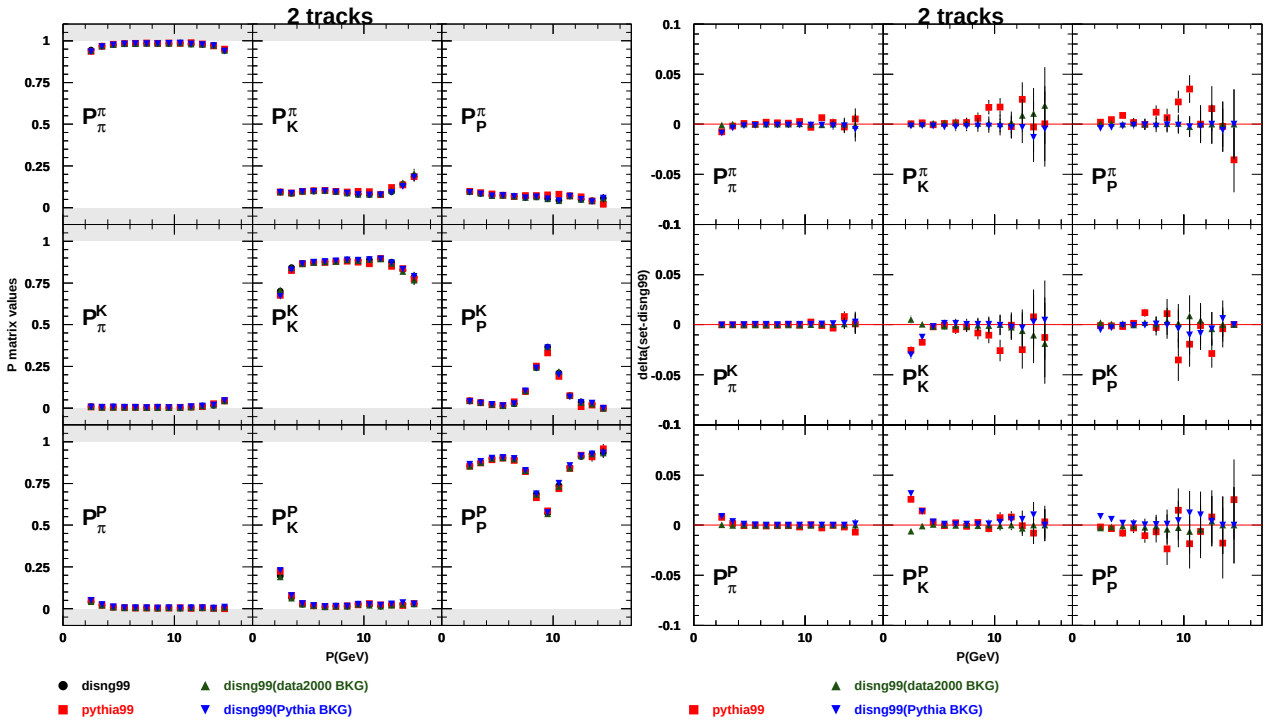

Figure 8: EVT Pmatrix sets (99 geometry, 2 tracks)

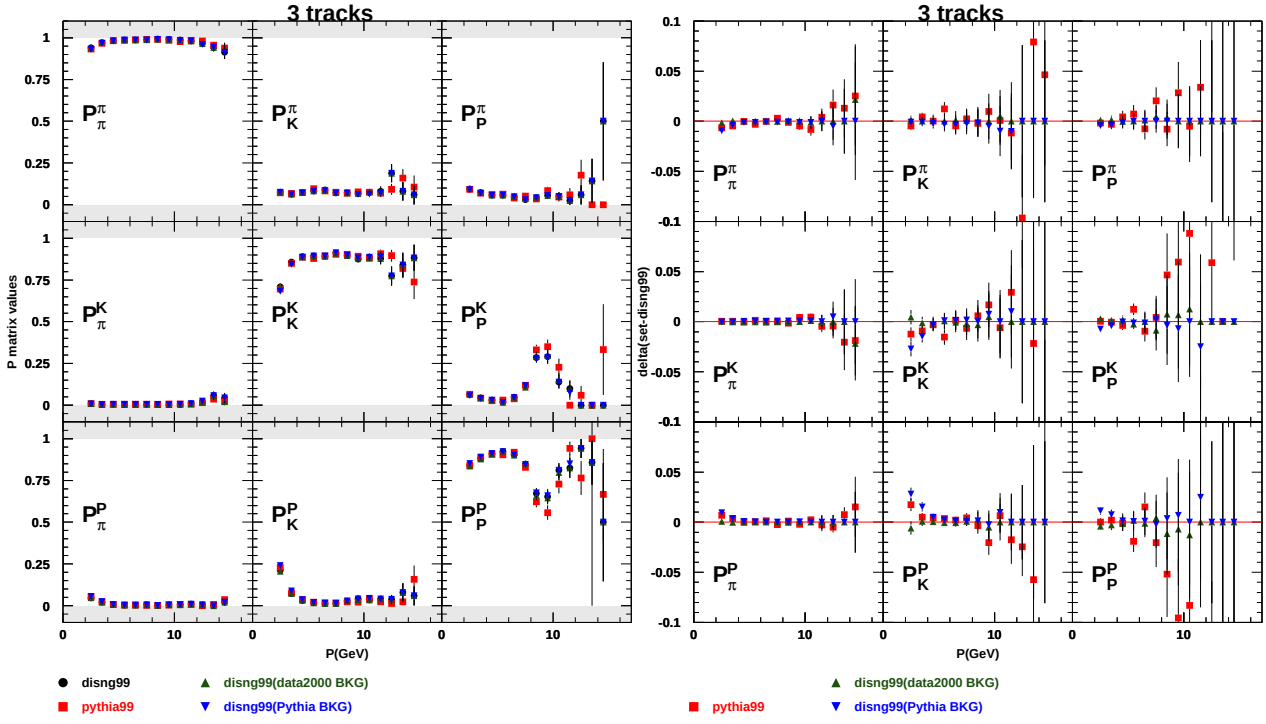

Figure 9: EVT Pmatrix sets (99 geometry, 3 tracks)

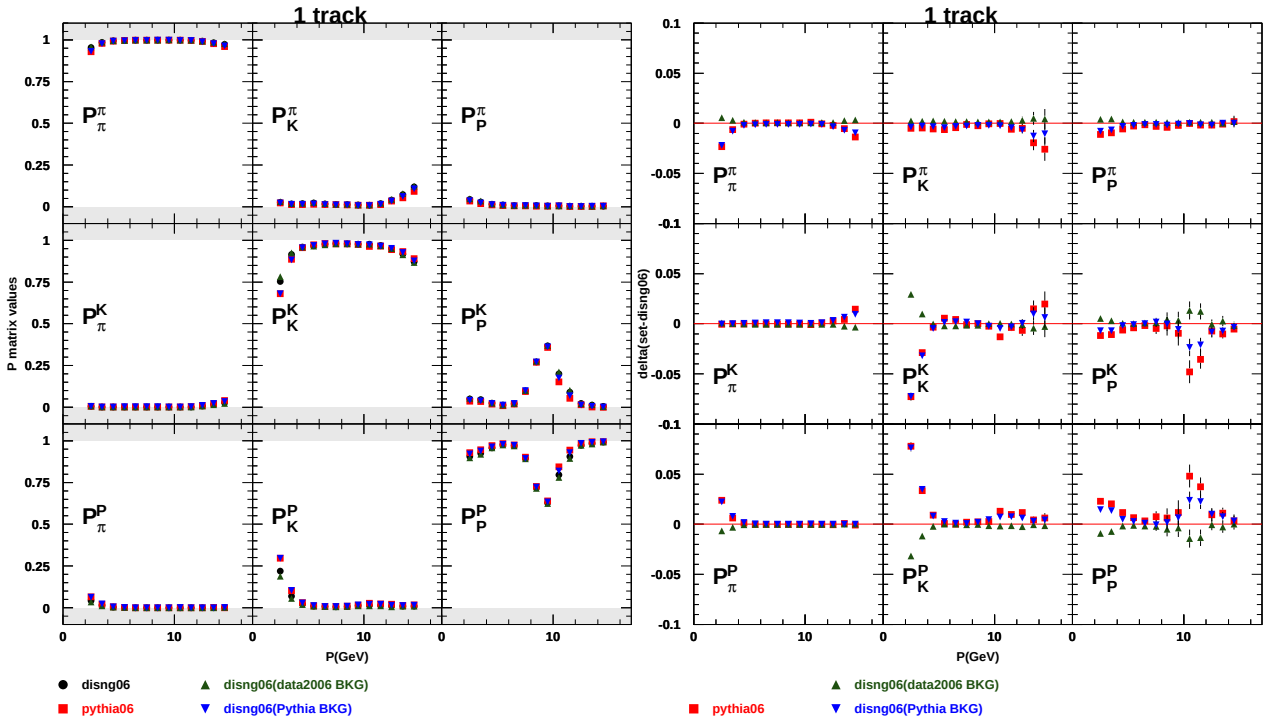

Figure 10: The four EVT Pmatrix sets for the 06 geometry and 1 track per detector half: disng06 is used for the central values, the systematic error is estimated using a different generator (pythia06) and disng99 with different background assumptions (disng06 with BKG from 2006 data and Pythia BKG, respectively). The plot on the right shows the difference of the Pmatrix values to disng06 (own BKG).

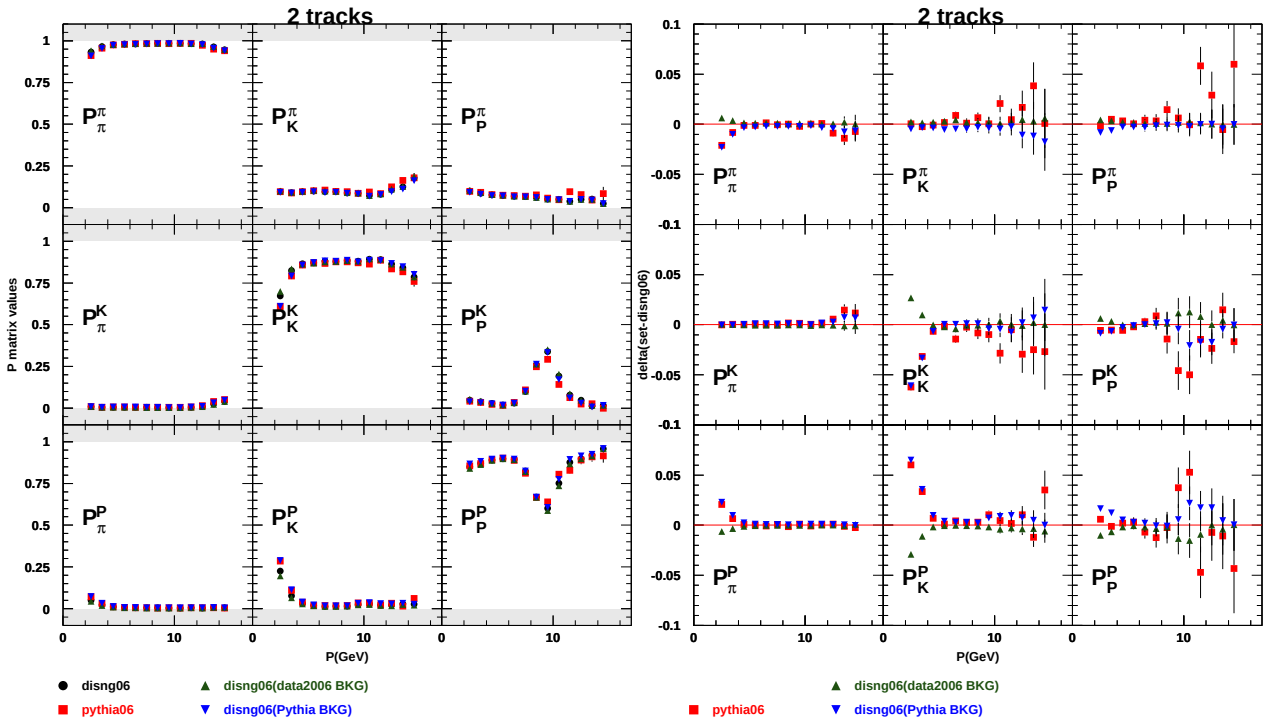

Figure 11: EVT Pmatrix sets (06 geometry, 2 tracks)

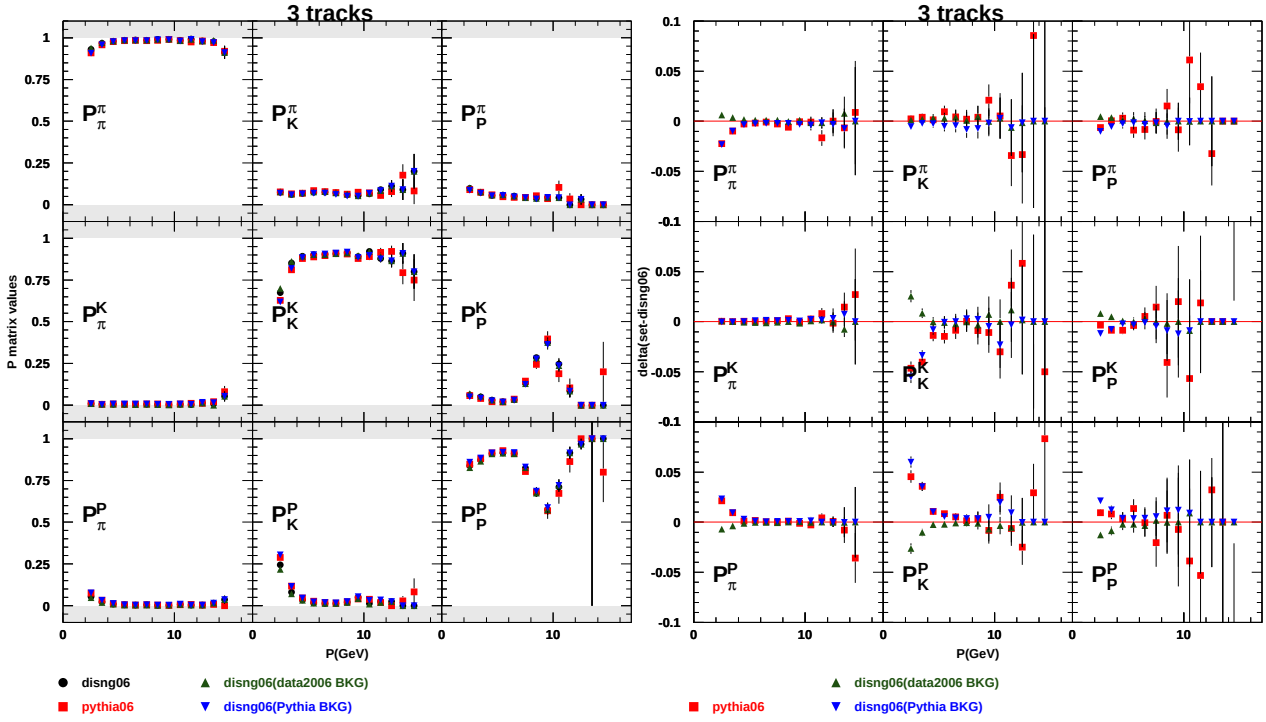

Figure 12: EVT Pmatrix sets (06 geometry, 3 tracks)

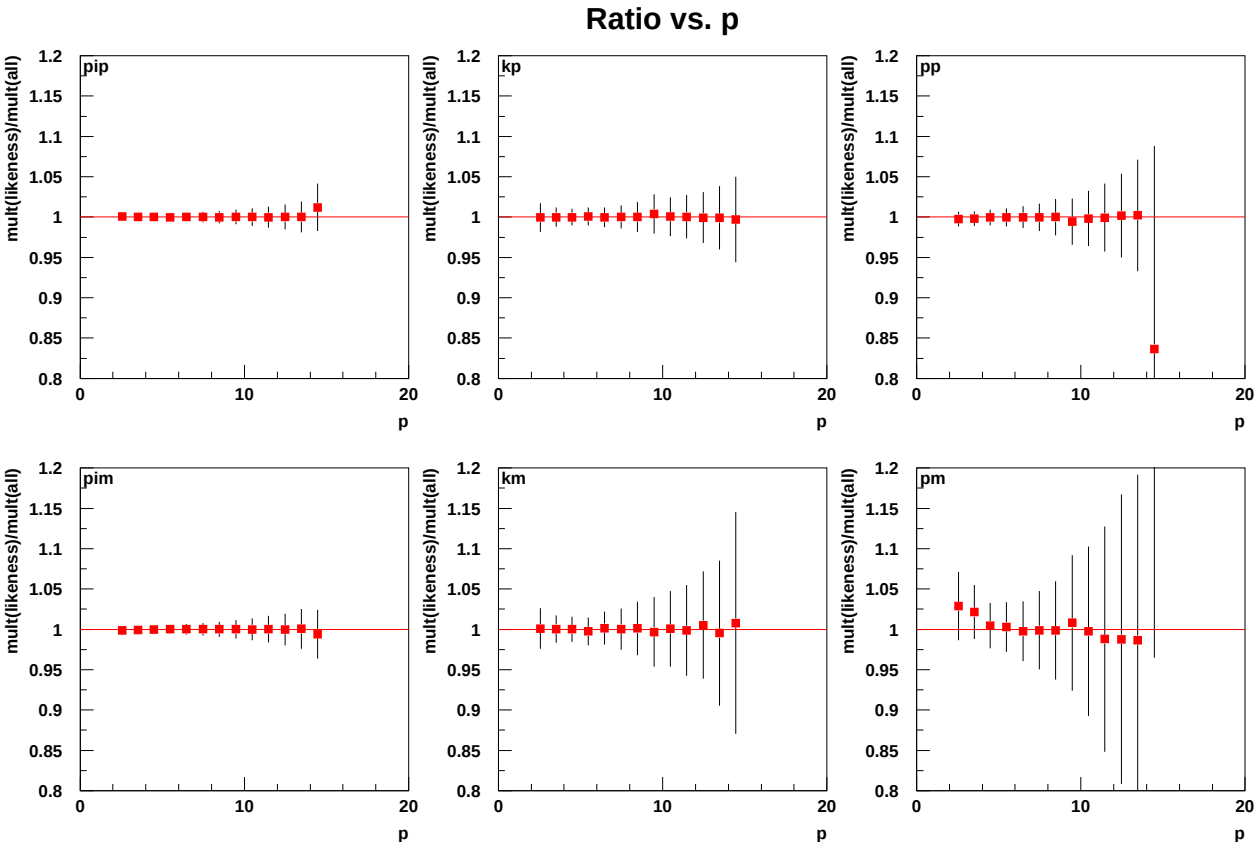

Figure 13: Ratio of multiplicities, unfolded with likeness separated Pmatrices (numerator) and charge combined Pmatrices (denominator), using the EVT PID method. The difference is marginal.

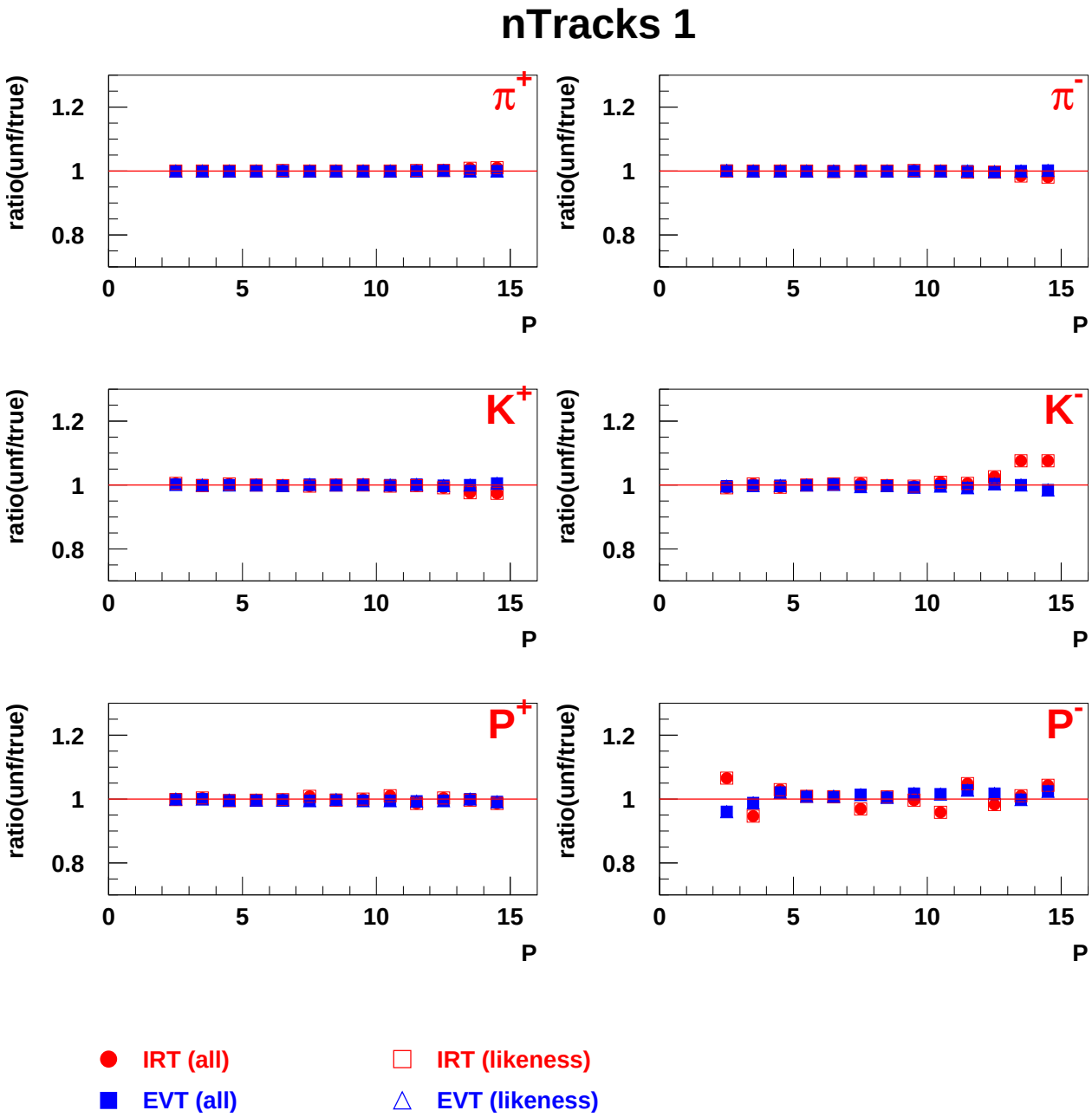

Figure 14: PEPSI challenge comparison for IRT and EVT. Shown is the ratio of unfolded to true hadron yields for **single track/(detector half)** hadrons. Both yields and Pmatrices are from the same MC production (disNG, 99 geometry, own BKG file).

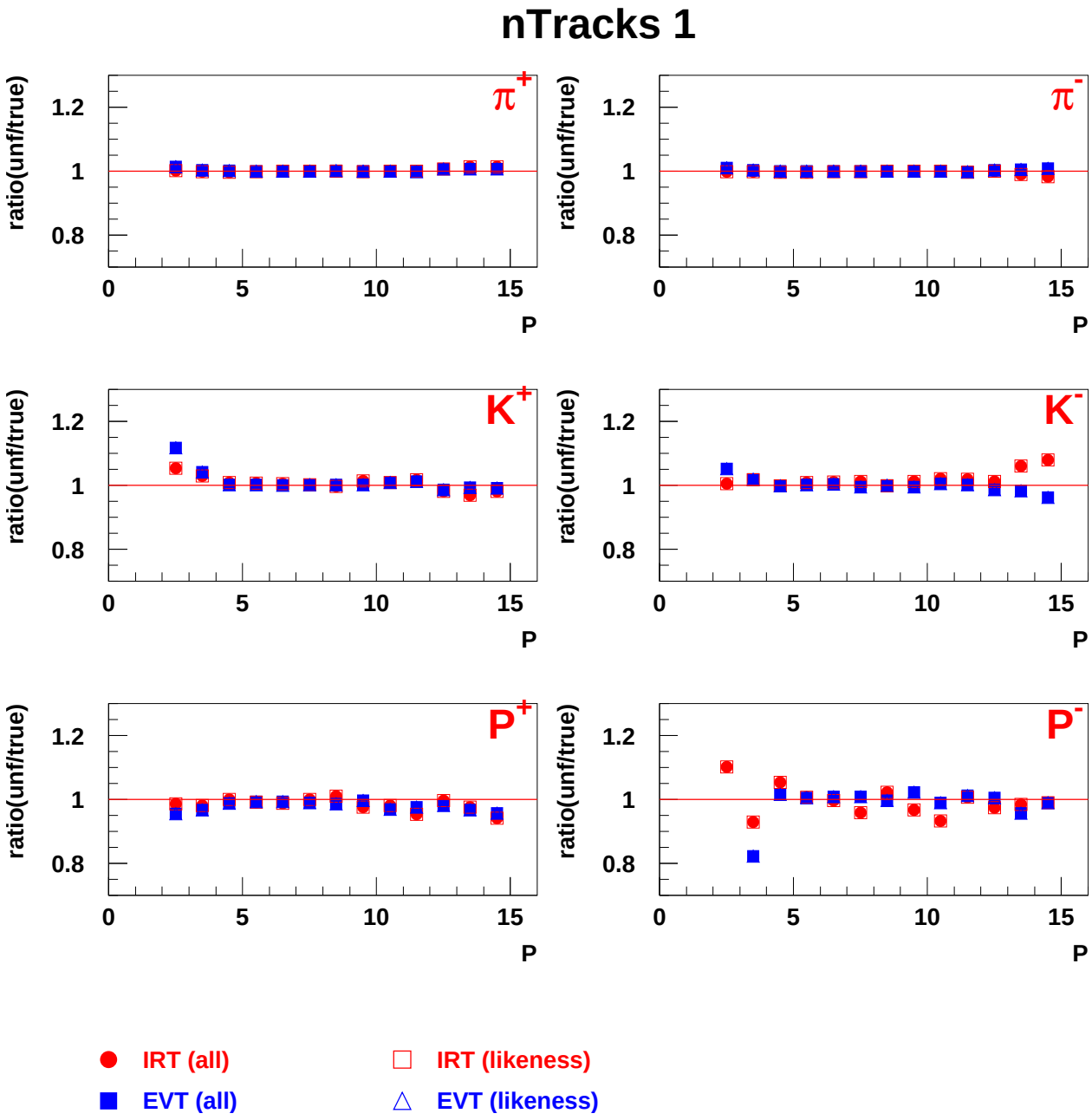

Figure 15: Same as Figure 14, but here the Pmatrices were extracted from a different MC production (pythia, 99 geometry, own BKG file). The hadron yields are the same as in Figure 14.

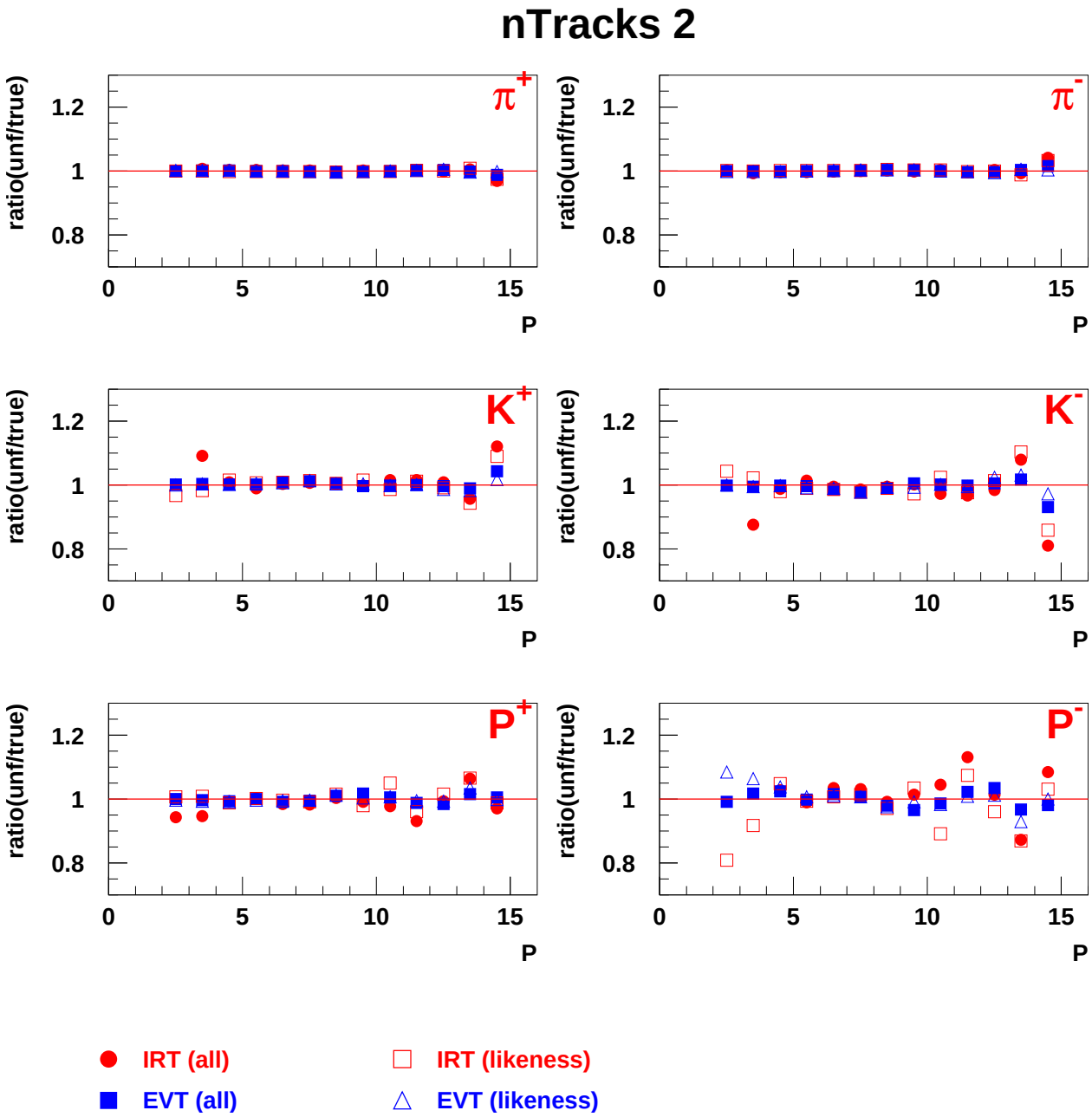

Figure 16: PEPSI challenge comparison for IRT and EVT. Shown is the ratio of unfolded to true hadron yields for **two track/(detector half)** hadrons. Both yields and Pmatrices are from the same MC production (disNG, 99 geometry, own BKG file).

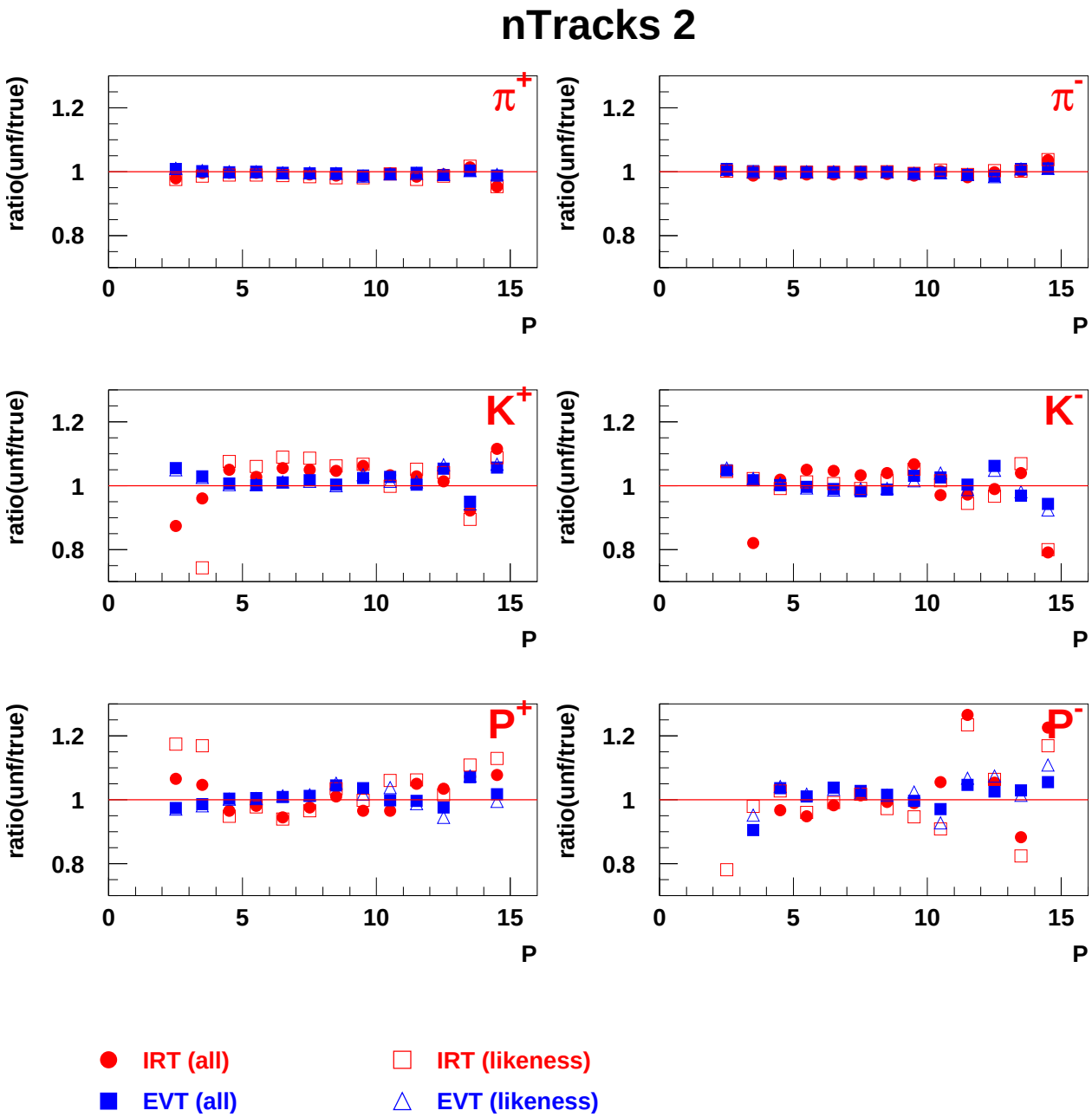

Figure 17: Same as Figure 16, but here the Pmatrices were extracted from a different MC production (pythia, 99 geometry, own BKG file). The hadron yields are the same as in Figure 16. Note that for K^- and anti-protons some data points at low momentum are actually off scale.

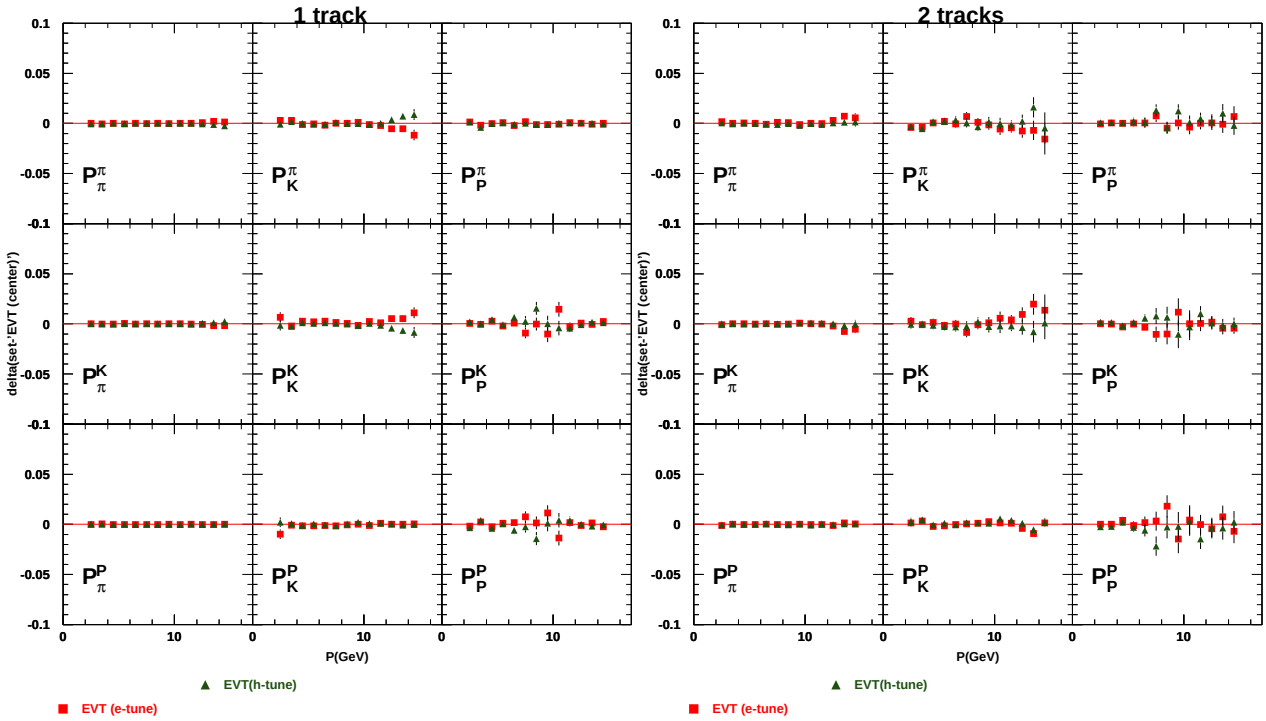

Figure 18: EVT Pmatrices extracted from three MC productions using the "center", "e-tune" and "h-tune" RICH MC settings, respectively. Shown is the difference of the e-tune and h-tune matrices to the center tune. The EVT method is much less influenced by the different RICH MC tunes than IRT (compare Figures 30 to 32)

option for EVT (and in principle DRT and IRT, although at the moment these algorithms do not calculate such a possibility) but is independent of the PID values from the other detectors (TRD, Preshower, Calorimeter). The input to the PID algorithm from the RICH (PID_{cer} in [2]) is always taken from IRT, regardless of what method the user selects from the smRICH table. HERMES lepton / hadron PID is unaffected by the addition of the EVT algorithm.

2.1 The weakness of DRT

The major disadvantage of the DRT algorithm is that it considers only a single track at a time. If, for example, there are two tracks in the same detector half, the PMT hits from both tracks are included in the observed hit pattern but only one track at a time is simulated. If the tracks and their corresponding rings in the RICH are well separated this means that the likelihood for all PTHs will decrease (compared to if there was only a single track), due to the high number of "unexpected" PMT hits coming from the other track. Since all the likelihoods decrease by the same amount this is not a problem as it does not change the most likely PTH. However, if the tracks are close and parts of the rings overlap, it is possible to misidentify PMT hits as belonging to the wrong track. For example, in Figure 19a, a proton (blue), which makes no ring of its own, falls between the C_4F_{10} and the aerogel rings of an electron (black) and the hits of the electron track allow the proton to be misidentified as a kaon (red). In Figure 19b the proton (blue) falls very close to the electron and the electron's C_4F_{10} ring (black) is identified as a kaon aerogel ring (red). In both cases the electron identification is unaffected but the proton is misidentified as a kaon, due to its proximity to the electron. Such an effect can be seen in the data by looking at the difference in x-position at the RICH between a hadron and a lepton, as in Figure 5.

Due to this inefficiency and its particularly strong effect on analyses where multiple tracks in a detector half are required, it was decided to implement an event level algorithm to better identify multi-track events.

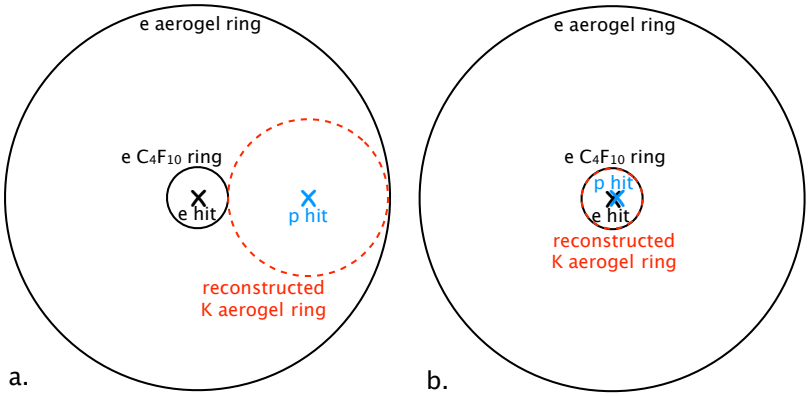

Figure 19: Examples of how tracks are misidentified when the track level RICH algorithm is used.

3 The EVT algorithm

In order to better identify tracks such as those described above, the EVT algorithm was devised. The idea is to evaluate all possible Combined Particle Type Hypotheses (CPTHs), combinations of PTHs for all the tracks in the detector half for one event. A CPTH is created by adding together the hit patterns for individual tracks from DRT and then evaluating the likelihood for the entire event topology. It is described in detail below.

3.1 Implementation

The EVT algorithm has been added into the RICH library, richlib. The EVT algorithm is called from the PID Scheduler function, *PSExec*. The main function, *event_Event*, performs both DRT and EVT. *event_Event* is only called if there are two to five tracks in the detector half. The upper limit on the number of tracks is set to constrain the amount of space used to store the array of CPTH likelihoods, which scales like $(\#\ of\ particle\ types)^{(\#\ of\ tracks)}$. In the case of a single track or more than five tracks, the existing DRT code (*drt_Track*) is called, not the modified DRT code that is within the *event_Event* function. It has been verified with a small test production that the changes made to the DRT code in the *event_Event* function have had no effect on the output of DRT to the μDSTs.

The *event_Event* function first performs the DRT algorithm on each track, storing the output

as well as the simulated PMT hit patterns. Then, the event level algorithm is run by cycling through each CPTH. For each CPTH, the expected fractional number of hits on a PMT from the relevant individual tracks ($Hits_{DRT}$) are added together and a background is added. The calculated hit pattern ($Hits_{hyp}$ with a non-integral number of "hits" in each PMT) is then compared to the observed hit pattern ($Hits_{obs}$ with 1 or 0 "hits" for each PMT) and the likelihood of this hypothesis ($L(CPTH)$) is computed. Mathematically, we have

$$Hits_{hyp}(\mathrm{CPTH}, N_{\mathrm{PMT}}) = \sum_{tracks} Hits_{\mathrm{DRT}}(\mathrm{PTH}, N_{\mathrm{PMT}}) + Background(N_{\mathrm{PMT}}) \quad (1)$$

$$P(\mathrm{CPTH}, N_{\mathrm{PMT}}) = 1 - e^{-Hits_{hyp}(\mathrm{CPTH}, N_{\mathrm{PMT}})} \quad (2)$$

$$L(\mathrm{CPTH}) = \sum_{N_{\mathrm{PMT}}} log\Bigg(\Big[P(\mathrm{CPTH}, N_{\mathrm{PMT}}) * Hits_{obs}\Big] \quad (3)$$

$$+\Big[(1 - P(\mathrm{CPTH}, N_{\mathrm{PMT}}) * (1 - Hits_{obs})\Big]\Bigg) \quad (4)$$

The likelihood of the current CPTH is compared to the previously noted most likely CPTH and the more likely of the two is stored as the current most likely CPTH. This is later used to extract the iType for each track. For each track, the likelihood of the current CPTH is also compared to the previously noted most likely CPTH with a different PTH for that track. The more likely of the two is stored as the most likely CPTH with this PTH for this track. This information is later used to calculate rQp. Finally, after all the CPTHs have been evaluated, the iType for each track is assigned according to the most likely CPTH, and the rQp for each track is computed, as described in Section 3.2.2. An example is described in Section 3.4. Details of improvements on the background values for both DRT and EVT can be found in section 4.

3.2 Quality parameters

In order to evaluate the quality of the particle identification for each event, both event level (G(1,2), G(1,3)) and track level (rQp) quality parameters were investigated.

3.2.1 G(1,2) and G(1,3)

G(1,2) is the log of the ratio between the top two most likely CPTH. G(1,3) is the log of the ratio of the most likely and third most likely CPTH. These quality parameters were suggested in [1]. In this simulation only events where the momentum of each track is greater than 2GeV were considered. However, many events contain a low momentum track, which does not emit many Čerenkov photons, along with other higher momentum tracks. For such an event, the CPTHs where the low momentum track is a kaon or a proton have the same likelihood. This means that such an event has a very low G(1,2) and G(1,3), even if all tracks except for the low momentum track are well identified. As a consequence, the correlation between G(1,2) and G(1,3) for properly identified and misidentified tracks shown in [1] disappears when events with any track momentum are considered. For this reason the quality parameters G(1,2) and G(1,3) were abandoned.

3.2.2 rQp

In order to evaluate the quality of identification for individual tracks, a track level rQp was devised. It is the log of the ratio between two CPTH likelihoods. The first is the most likely CPTH. The second is the next most likely CPTH in which the track in question was hypothesized as a different particle type as in the most likely CPTH. For an example see Section 3.4.

3.3 Output

The EVT algorithm provides only additional output in the μDST files. The only change from previous productions is that DRT is now run for all tracks and uses the same non-constant background file as EVT, as described in section 4.1. The BEST link is still evaluated in the same way and never links to the EVT output. The additions to the μDST tables are listed below.

3.3.1 The g1Track table

A new link, EVT, was added to the g1Track table. This is similar to the preexisting DRT, IRT, and BEST links. For each track where the event level algorithm is evaluated, EVT links to a (different) row in the smRICH table. Since EVT is only for events with multiple tracks in a detector half, events with a single track in the detector half have no EVT information. In this case the EVT link was set to various values, depending on the data production. See section 3.6 for a listing of avaiable data production and the status of the EVT single track link. For all future productions the EVT single track link will be set to DRT. See section 1.2 for the recommend smRICH links to use, and section 3.5 for pseudo code.

3.3.2 The smRICH table

iMethod: This is set to 3 for the EVT method (IRT is 1 and DRT is 2).

iType: This is set to 3, 4, or 5, corresponding to pion, kaon, and proton respectively. Although the EVT algorithm considers electrons (positrons) in its CPTHs, it is never returned (Please see note in section 2 regarding lepton / hadron PID and the electron hypothesis) . This is due to Monte Carlo studies which showed that the number of misidentified tracks (not counting electron / pion misidentifications) is minimized when the electron hypothesis is considered, but that most tracks identified as electrons were in fact pions. So, all CPTH are considered and the one with the highest likelihood is chosen. Then, each track is assigned iType according to the content of this CPTH. If the electron hypothesis is the most likely for a given track, iType is set to 3 for that track.

rQp: When calculating rQp, the electron hypothesis is never considered. If rQp is zero, iType is also set to zero. See section 3.2.2.

rProb[5]: This contains the "row sums" for each particle type hypothesis: 1=electron, 2=muon, 3=pion, 4=kaon, 5=proton. The "row sum" is obtained by computing the sum of the logs of the likelihood where the track in question is hypothesized to be the given particle type. As the muon hypothesis is never considered, rProb[1] is always set to 0. For an example see section 3.4.

All other columns of the smRICH table are set to a default value of -9999.

3.4 An Example

A hypothetical matrix of CPTH log-likelihoods is shown in Table 2. In this case there are only two tracks in the detector half. The most likely CPTH is track1 = kaon, track2 = pion, with a log-likelihood of -24. iType is then track1.iType=4, track2.iType=3.

rQp for track1 is $(-24) - (-32) = 8$. This is the difference of the two most likely CPTH log-likelihoods (or the log of the ratio of the likelihoods) where track1 has different PTHs, without considering the electron hypothesis. For track2, rQp is (-24) - (-35) = 9. Here the confidence that this track is a pion is reflected by the high rQp, and it is not artificially lowered by the relative unconfidence in track 1.

The rProb values for each track are filled with the values in the "row sum" column / row. These are just the sum of the likelihoods in that row / column.

		Track 1				
		Electron	Pion	Kaon	Proton	"row sum"
	Electron	-37	-35	-30	-41	-143
	Pion	-32	-32	-24	-37	-125
Track 2	Kaon	-48	-47	-41	-35	-171
	Proton	-51	-53	-48	-71	-223
	"row sum"	-168	-167	-143	-184	

Table 2: An example of the log-likelihood matrix for an event with two tracks in the detector half.

3.5 How To Use the EVT algorithm

Analyzers can use the EVT algorithm information just as IRT, DRT and BEST were previously invoked:

$NATREL(g1Track, g1Track.EVT, smRICH, ok);$

In some productions (see notes on productions in the next section, 3.6) the EVT link for single tracks doesn't exist or links to BEST. If you want to link to DRT for single tracks (which is recommended), the following pseudocode is appropriate:

```
...
ntop = number of long tracks in the top detector half
nbot = number of long tracks in the bottom detector half
...
if( this is a top track ){
  nhalf = ntop;
} else {
  nhalf = nbot;
}
if(nhalf > 1){
  NATREL(g1Track,g1Track.EVT,smRICH,ok);
} else {
  NATREL(g1Track,g1Track.DRT,smRICH,ok);
}
if( ok ){
  itype = smRICH.iType;
} else {
  itype = -999;
}
```

or alternatively

```
NATREL(g1Track, g1Track.EVT, smRICH, ok);
if (ok && smRICH.iMethod==3){
  // If this is really EVT (method=3) then take this information
  itype=smRICH.iType;
} else {
  // if there is no EVT information (single track OR EVT failed), get DRT information
```

```
  NATREL(g1Track, g1Track.DRT, smRICH, ok);
  if (ok){
    itype = smRICH.iType;
  } else {
    itype = -999;
  }
}
```

3.6 Current Data and MC productions

At the time of writing not all data years have been reproduced to include EVT. The current status and some comments on existing productions follows.

Productions with EVT:
00d2 - EVT single tracks link to BEST, so one must manually link to DRT to single tracks. See sample code in section 3.5.
04c1 - EVT single tracks link to BEST, so one must manually link to DRT to single tracks. See sample code in section 3.5.
05c1 - EVT single tracks link to BEST, so one must manually link to DRT to single tracks. See sample code in section 3.5.
06d0 - EVT single tracks link to DRT
07b1 (not yet released) - EVT single tracks link to DRT

All other data years have not yet been produced with EVT, but when they are EVT single tracks will link to DRT.

Several MC productions exist with a variety of possibilities for the single track EVT link. Some have a NULL link, some link to BEST and some link to DRT. The productions in
`/mcdata06/DATA/RICH_SYSTEMATIC_STUDIES/`
and any more recent productions should link to DRT for single tracks. The best way to check is to look at a few events in pb and compare the links in the g1Track table.

Up to date information can be found on the "RICH PID" page of the HERMES wiki

(http://hermes-wiki.desy.de/index.php/RICH_PID).

4 Further improvements to DRT and EVT

Both the EVT and DRT algorithms have been further improved by the inclusion of detailed information on each PMT: if it is hot, dead, or neither and what background hit level is expected. This information can be extracted from any HRC sample. A single background file has been extracted for each year that has been used in the μDST production of that year's data.

4.1 Background

In the DRT and EVT code, a constant 0.0001 was added to the expected number of hits for each PMT (see Equation 1). However, this can be optimized by adding an appropriate (different) value for each PMT. This has been implemented in the code and the background values for each PMT are read in from an external file. The background numbers in this file are obtained from the data or MC by adding up hits in the RICH PMTs when there are no tracks in that detector half and normalizing to the number of these "empty" events.

In MC one must also require that at least one track is seen in the other detector half to better simulate a "trigger" event in data. In the data, several variations were investigated, including allowing a short track in the "empty" detector half, requiring and not requiring a full track in the other detector half, and requiring a DIS lepton in the other detector half. Among the stable backgrounds (see section 4.1.1) these variations had no significant effect on the background file.

Two typical background files (from the 2000 polarized data and the 2007 data) are displayed in Figure 23. The PMTs are numbered down columns of 26. Thus, the small periodic spikes results from more hits occurring in PMTs near the (top-bottom) center of the detector. The overall increase near 1000 corresponds to an increase in hits in the (left-right) center. Thus, the overall shape of the background displays the "banana" shape familiar from the onsite plots. The negative values indicate dead tubes, as discussed in section 4.2. Ratios of background from different years appear in Figure 25. Projected into the tube map plane, this translates to the typical "banana" shape well

known from the online data quality plots, so the background is highest in the central region of the RICH detector. This shows that the background is mostly not due to electronic noise and similar effects (which were supposed to be taken into account by the flat background). Instead, most of the background is coming from untracked particles passing through the RICH detector. This is further evident when looking at the RICH event display (Figure 20): often, the background is not uncorrelated but comes in the form of ring patterns.

4.1.1 Data background files

It was found that the background files extracted from some subsets of data display a step function where one half of the detector (both top and bottom) has significantly higher background than the other. For example, Figure 24(left) shows the background extracted from high density data taking in 2000; the polarized data set in 2000 (Figure 23left) displays no such jump in the background values. This is the most severe example, but a similar effect is also seen in the 2006 data both during high density data taking and in all data taken with the recoil magnet on; the recoil magnet off data does not display this jump. It is supposed that this effect may be related to the lack of/presence of a transverse target magnet, but it has not been studied further. Nonetheless, in order to avoid possibly introducing a (small) left-right asymmetry in the data, in all cases a background file with no such step function will be used in the data productions. The background file extract from the 99 and 04polarized data displayed different strange behavior, in these cases showing "noisy" results (see Figure 24(right)). Such files were also not used for data productions. For all data years runs with RICH problems (as noted in the logbook) were eliminated. In some cases other limits to the HRC run list were imposed. They are explained here.

98 - All data used (only polarized data produced a file that was 20% smaller)
99 - 00 file used and dead/hot tubes from 98 added. (99 data produced "noisy" results)
00 - Polarized data used. (HD data contains "step function")
02 - Unable to produce file at this point due to unavailability of HRC files.
03 - Unable to produce file at this point due to unavailability of HRC files.
04 - HD data used (Polarized data produced "noisy" results)
05 - HD data used (Polarized data produced difference between top and bottom, see Figure 29)
06 - Recoil Magnet Off data used (Recoil on and HD data both contain "step function")

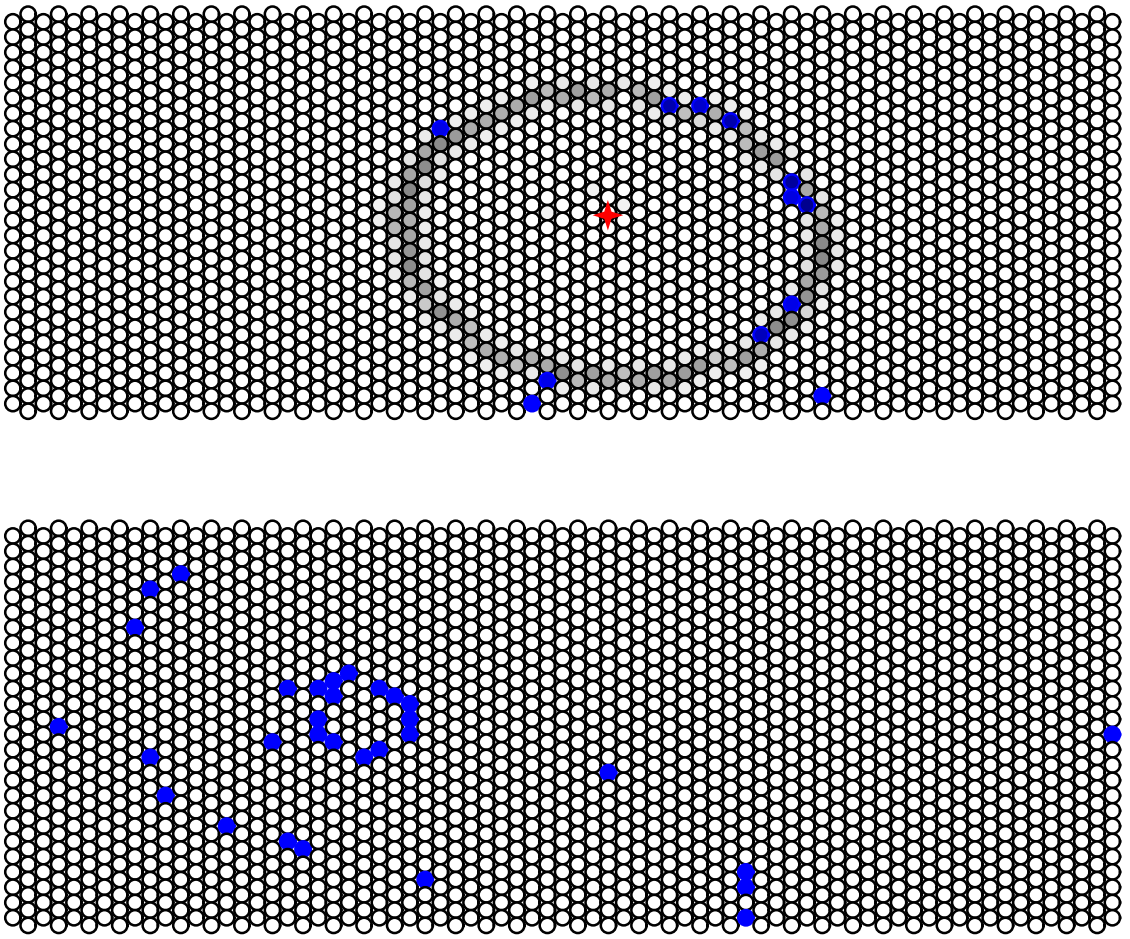

Figure 20: The HERMES RICH Event display (HeRE) showing Event 57596322 from the 2005 data. In the top half of the detector the PMT hits (in blue) that surround a track (in red) are identified as a pion ring (in gray). In the bottom detector half there are no tracks but a clear pair of rings is present.

07 - All data together used (HD data is slight higher, Recoil On data is slightly lower, Recoil Off data has low statistics)

Background files for all data productions can be found on the PCfarm, at `/group01/richgrp/BackgroundFiles`

4.2 Dead tubes and Hot tubes

Dead tubes, if not accounted for, have only a small influence on the likelihood. However, they can unnecessarily decrease the likelihood of a particular hypothesis that expects a dead tube to fire. Simply zeroing the number of expected hits (the decimal number mentioned above as $Hits_{hyp}$) in such tubes is also insufficient as this can lead to a large decrease in the likelihood if the tube does fire. Thus, the best course of action is to skip over this tube entirely when computing the likelihood, making a small hole in the detector. Such tubes are known to the algorithm via the background files mentioned above. Dead tubes are denoted in the background file by setting the background values for this tube to -1. Dead tubes were identified by looking at the total number of hits in each PMT for all events with at least one track in the detector half. Each tube is then compared to the average number of hits over all PMTs. Any tube with 100 times less hits than average is considered dead. Tubes included in a dead tube list obtained from Elke were also marked as dead.

Hot tubes can have a similar small influence on the likelihood. Like dead tubes, if we now suppose that a hot tube will always fire and modify the expected number of counts accordingly, this can have a negative effect on the likelihood when the tube does not fire. Again the best solution is to skip over such tubes. Such tubes are noted in the background file by a value of -2. Hot tubes were identified "by hand" by examining a plot of the background values, such as Figure 23. Any tube that appears above the "grass" of the surrounding tubes is marked as a hot tube.

Both the EVT and DRT algorithms have been modified to skip tubes with background values less than 0 when computing the likelihood of a hypothesis.

5 Preliminary Monte Carlo Studies

Several Monte Carlo studies have been done to determine the effect of using different background files and hot and dead tubes. More recent studies done in determining the Pmatrix (see section 6) have called these preliminary studies into question, but they are included here for completeness.

For these studies the differences between the productions were inspected by comparing the Pmatrices as well as the F parameter which is discussed in [3] and is calculated as

$$F = \frac{P_\pi^\pi + 2P_K^K + P_p^p + 2Q_\pi^\pi + 4Q_K^K + 2Q_p^p}{12 * 15} \tag{5}$$

Where the P's indicate the efficiencies for each hadron type and the Q's indicate the corresponding purities. The denominator is simply to normalize the value to 1 when summed over all 15 momentum bins.

5.1 Background studies

Several different background files were used as input in three different MC HRC productions. The results are discussed below.

5.1.1 Beam pipe simulation ON and OFF

A small pythia MC production of K_s particles (using a selector) was run with the "dobeampipe" options switched ON and OFF. This option, when ON, sets the TraPipe parameter to YES and Enemin to 0.0001. When OFF, TraPipe is set to NO and Enemin is set to 0.0100. The initial purpose of this study was to evaluate if the dobeampipe option needed to be ON in order to accurately reproduce the background observed in data. The background files extracted from these two productions were very similar, with the ON production displaying only slightly larger background values. They are both shown in Figure 26 (note difference in scale compared to other background files shown). Both background files were significantly larger than that extracted from data (see Figure 23) due to the larger number of $K_s \to \pi^0\pi^0 \to \gamma\gamma\gamma\gamma$ events which decay downstream and

thus produce no tracks but do produce hits in the RICH when the photons pair produce. Since running the MC with the ON option takes both more time and more disk space and does not reflect the data background any more accurately than the OFF option, it was decided that this option should be OFF when producing the MC files used to extract Pmatrices.

Note: These background files were extracted before it was realized that one should not consider events where both detector halves are empty when extracting background. This change reduces the backgrounds from a peak value of 0.05 to about 0.03. Nonetheless, these studies show that the background from beam pipe ON and beam pipe OFF are very similar to each other and much larger than data, due to the all-K_s event sample.

5.1.2 The influence of different background files

Both the two K_s MC HRC file samples (ON and OFF) discussed above and a disNG MC HRC file sample were used in combination with 5 different background files to produce 20 different μDST productions. The 5 background files were:

"Zero" - a constant background value of 0.000001

"Flat" - a constant background value of 0.0001 (the value used before this upgrade)

"2000" - the background extracted from the 2000 data (both Pol and HD) but with the dead and hot tubes "smoothed" by averaging them with the surrounding tubes. Note that this background has a "step function" (discussed in section 4.1.1). See Figure 26(bottom left).

"disNG" - the (small) background extracted from the disNG MC sample. See Figure 26(bottom right)

"ON/OFF" - the (large) background extracted from the pythia K_s MC sample with the dobeampipe option OFF was used for both the OFF and disNG productions. For the ON MC sample the ON background was used. See Figure 26. Thus there was one μDST production of each different HRC productions (ON, OFF, and disNG) that was produced with the background extracted from that production.

The resulting F values for each of these productions can be seen in Figure 21. The main conclusion to draw is the algorithm is relatively insensitive to the choice of background file. The worst results are for the "zero" and "flat" files, which contain the least information. The pythia K_s productions (both ON and OFF) are a strange event sample and display some odd results (for example, sometimes the DRT algorithm has a higher F value than EVT), but again show that the algorithm is mostly insensitive to the choice of background file.

NOTE: This conclusion was called into question when differences in the Pmatrices were seen when variation were made to the background files. In these case a more "reasonable" MC sample was used which did not include any selector files. See section 6.4.

The final conclusions from this study are that for data the background file averaged over the whole year is sufficient and no smaller time periods need to be considered. For more discussion of the step function in the 2000 background files see section 4.1.1. For MC, if it is a simple DIS sample, such as the disNG production used in this study, the background file from the production itself can be used. Since extracting the background requires first producing HRC files and this is often inconvenient in MC productions, a standard background file extracted from a DIS MC sample may also be used. For MC productions such as the pythia K_s MC, where a selector has been used, the standard background files should be used since this "average" treatment is the most similar to how the data is treated.

5.2 Dead and Hot tubes studies

Further studies were performed with the disNG production mentioned above to see the effect of eliminating tubes. In the most extreme test, 50 tubes in each detector half were chosen as dead tubes. Included in the 50 were all of the tubes found to be dead in the 2000 data, the rest were randomly chosen. In the MC the response from these tubes was always set to zero. Two productions were run with this degraded data - one which used a background file with no dead tubes and one which included all the dead tubes and thus skipped over these PMTs when calculating the likelihood. Virtually no change was seen in the Pmatrices or the F values (see Figure 22). This is not surprising as the likelihood that an individual PMT will fire for any event is small even if a ring falls on the PMT, and to change the Pmatrices or F value not only the likelihood but also

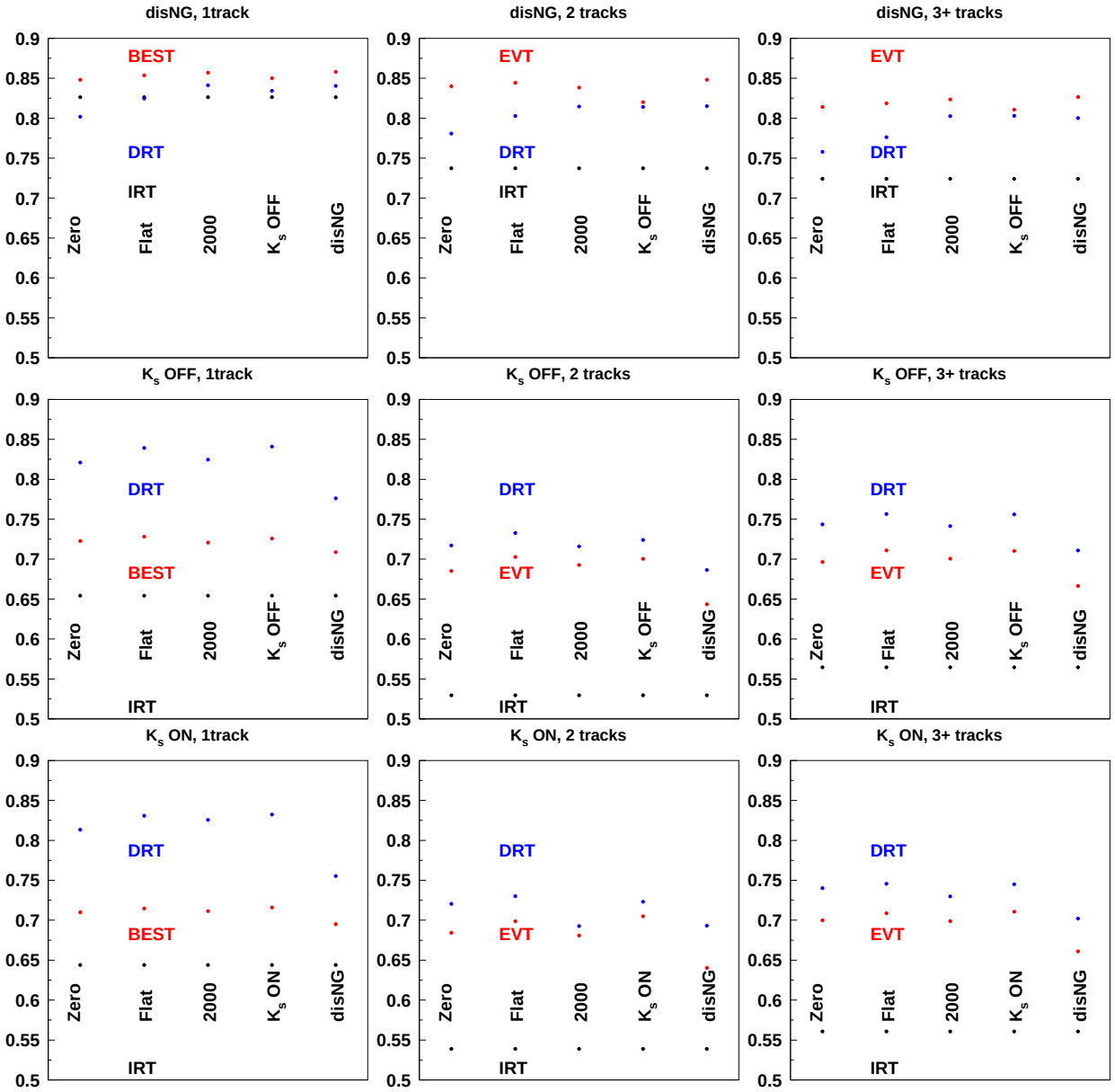

Figure 21: F values for (red) EVT, (blue) DRT, and (black) IRT. The columns refer to 1-track, 2-track, and 3+ track events respectively. The upper set of plots refers to the disNG MC sample. The middle row is the K_s OFF sample and the bottom pair is the K_s ON sample. The x-axis denoted the background file used: Zero, Flat, 2000, K_s ON or OFF, and disNG. As IRT does not make reference to the background file its value is constant with respect to the different backgrounds.

the final particle type decision must be changed. However this test is useful in demonstrating that holes in the detector from dead tubes, as well as the holes we create by ignoring output from hot tubes, should not significantly degrade our particle identification ability.

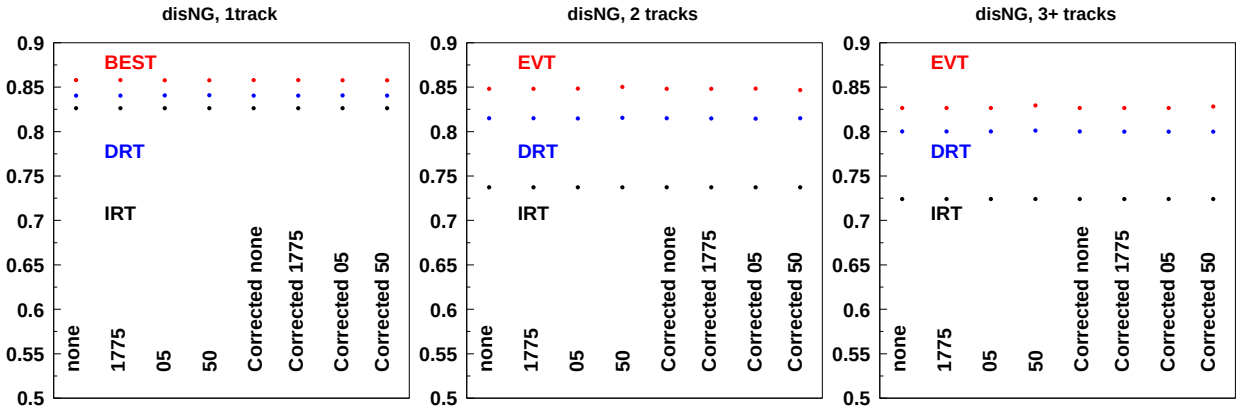

Figure 22: F values for various configurations. The panels denote 1, 2, or 3+ track events. The colors denote the RICH method. Along the x axis, the first four values show the F value for: no dead tubes, tube 1775 top dead, the 05 dead tubes dead, and 50 tubes dead, respectively. The second four points show the effect of again forcing no output from these tubes but now they are "corrected" for by skipping over these tubes when computing the particle type likelihoods. In IRT there is no correction made so the printed values are unchanged from the first four points.

No direct tests of hot tubes were done. However, the dead tubes test shows that ignoring these tubes is not harmful to our particle identification, especially as there are only 5-10 hot tubes, as can be seen in Figure 23.

A study was done where the data background file was read in and the values were used as the probability of a random "noise" hit. It was later realized that this is an inappropriate test of the effect of background because the MC already has physics background in it which comes from real particles (such as $\pi^0 \rightarrow \gamma\gamma \rightarrow e^+e^-e^+e^-$). The background is then not distributed evenly between all events (as noise would be) but rather is collected in rings and thus affects fewer events. However, this is a good (although extreme) test of hot tubes which do fire independently and randomly. In this test it was shown that such random hits reduce the identification quality. Thus, as this shows hot tubes may be harmful, and the previous study shows ignoring them is not harmful, we conclude that skipping these tubes in calculating the likelihood is the best option.

6 Pmatrices and Systematic Error

First, a brief history of Pmatrices is given. In order to try to determine the systematic uncertainty of the Pmatrices several different kinds of MC productions were compared. These are described in sections 6.4-6.6. Different types of charge separation are explored in section 6.3. The proton Pmatrices were also calculating from data using lambdas, this is discussed in section 6.7. Finally, a prescription for how to find the systematic errors is given at the end of this section.

6.1 History: The version 3.0 Pmatrices

The previous Pmatrices were tuned with and computed exclusively with the IRT method. The electron, hadron and center tuned IRT Pmatrices are shown in Figures 30 to 32. During the course of these studies we tried to replicate the version 3.0 Pmatrices found at

`/group01/richgrp/dPsys/pmatrix_v3.0`

but were unable to reproduce them, despite using various MC samples (including the one believed to be used to first produce these matrices). This is shown in Figure 33 for the case of a single track in the respective detector half. Here, *v3.0* refers to the version 3.0 Pmatrix, *IRT (TTech)* is a new extraction from the MC production which is believed to be originally used for the v3.0 matrices, and *IRT (RMC)* are IRT matrices from a new MC production (disNG with 1999 geometry and a selector file to enhance high momentum) which is also used to extract the EVT Pmatrices. The current Pmatrix extraction code was crosschecked by Achim and Rebecca. The variation between the old and new IRT Pmatrices is small but significant, particularly at low momentum for kaons and (anti)protons (see Figure 68). Old IRT is also compared to the most recent version in Figures 69 and 70.

6.2 Cuts

After substantial discussion of if cuts (fiducial volume cuts, DIS cuts, etc.) should be applied to the Pmatrices it was decided that only a very minimal set of cuts should be applied (see also Table

3).

Tracks are required to be "long tracks" (ie to consist of a front and a back partial track). They need to satisfy the usual fiducial cuts in terms of position at the calorimeter and the spectrometer magnet field clamps. The lepton z vertex is required to lie within the target cell, while for hadrons the range is extended to 100 cm to allow for decaying particles. There is no constraint on the vertex distance.

Events need to have a scattered lepton in acceptance. No further kinematic cuts are applied.

These same event level cuts are also applied when calculating the background for both MC and data.

Table 3: Cuts used to create standard Pmatrices

Event level cuts:	require beam lepton
Track level cuts:	RICH EVT link exists ($\Rightarrow$ long track)
	fiducial volume cuts:
	$\rightarrow$ z vertex cut
	$\rightarrow$ calo position
	$\rightarrow$ field clamp acceptance cuts

6.3 Charge and Likeness separated matrices

In an effort to improve the Pmatrices, charge separated Pmatrices were constructed. While the difference between the two charges were small, when the ratio of the charge separated to charge combined matrices was taken the difference becomes more apparent. See Figures 35 to 37 for IRT and 38 to 40 for EVT. These small differences make a large effect for IRT in regions with small fluxes - particularly for protons / antiprotons, as can be seen in Figures 71 and 72. The difference for EVT however is very small, as can be seen in Figures 73 and 74. Unfortunately, making charge separated (positive and negative) Pmatrices is impractical since, given the dominance of the beam in our data, the Pmatrices are beam charge dependent. So, we created "likeness" Pmatrices. They are defined as:

like = the track in question is accompanied by at least one like-charged track in the detector half

unlike = the track in question is the only track of this charge in the detector half
With this definition the charge of the beam becomes irrelevant since only the relative charge of tracks is considered. To be explicit, the following combinations, where the first sign is the track in question, are like / unlike:
like: (++) (- -) (++ -) (- - +) (+++) (- - -)
unlike: (+) (-) (+ -) (- +) (+ - -) (- ++)

In a charge separated, momentum dependent PEPSI challenge the charge separated Pmatrices give identical results to perfect iLType PID. This is merely an indication that the Pmatrices and RICH unfolding code work correctly. The likeness and charge combined Pmatrices do not exactly produce the correct yields, but they are close. See the RICH yields divided by the true yield ratio for all kinds of PEPSI challenges in Figures 49 to 66, which are discussed in more detail below.

6.4 Background Study

As more detailed (and higher statistic) studies were done it was supposed that the background file used as input to DRT and EVT may have a significant effect on the identification and therefore Pmatrices. In order to study the effect, the RICH EVT algorithm was rerun on one MC production (disNG, 2006 geometry), using several different background files. The result was six different MC data sets, which were identical except for the hadron PID of the EVT algorithm.

The following background files were used:

- **disNG06:** extracted from the disNG 2006 MC production itself. Thus this background assumption reflects the best knowledge about the background affecting the RICH in the used MC. This data set was used as baseline to compare the other background assumptions to. (Figure 28 left)
- **Pythia99:** extracted from a Pythia production using the 1999 geometry (Figure 27 right)
- **polData05:** extracted from polarized 2005 data (Figure 29)
- **disNG99:** extracted from a disNG MC production using the 1999 geometry (Figure 27 left)
- **polData00:** extracted from 2000 data (Figure 23 left)

- **flat BKG:** old flat (=0.0001) background used so far for the DRT algorithm

The backgrounds given above can be divided into three groups. In the first three cases, the background is comparatively large, up to values close to 0.02. The backgrounds from the 2000 data and the disNG background using the same geometry show a similar shape, but stay below 0.01. The flat and almost negligible "flat BKG" forms the third group. The first two groups span the variation in background size observed in the data (see Figure 25).

The effect of these three background levels on the Pmatrices can be seen in Figure 41 on page 64 and Figure 42 on page 65. Here, the Pmatrices from the various background assumptions are shown as a ratio to using the background extracted from the MC production itself. Clearly, a lower background assumption generally leads to less misidentifications of lighter hadrons as heavier ones. At the same time, less assumed background[2] increases the likelihood to misidentify kaons and protons as pions (and, to a lesser extent, protons as kaons). The effect is stronger for single tracks compared to tracks having a partner in its detector hemisphere. This effect can be understood if you assume that with a large background file hits are given less value, which means the proton hypothesis (aka the hypothesis that there is in fact not a ring present) increases. This is true for both the case of the particle actually being a proton, where hits due to other sources are not identified as belonging to the proton track, and in the pion hypothesis, which hits belonging to the pion track are not identified as belonging to the track. Thus, a background file with large values will tend to favor protons, giving better values in the upper right half of the Pmatrix and worse values in the lower left half of the matrix. Conversely, a lower background assumption will favor pions and give the opposite results.

Figure 75 shows multiplicities obtained from unpolarized proton data, using the 00d2 production and the Pmatrices extracted using the different background assumptions. As is to be expected, disNG06 BKG, Pythia99 BKG and PolData05 BKG yield very similar results (except for antiprotons at low p), the lower backgrounds disNG99 BKG and PolData00 BKG show a stronger deviation, which gets more extreme for the flat BKG (see also the ratios in Figure 76).

[2]It is important to keep in mind that we are not varying the background itself, just the background *assumed* in the EVT algorithm.

6.5 Physics Generators Study: disNG vs. Pythia

Pmatrices were generated from MC using the disNG and Pythia generators. For each production the background file extracted from the production itself ("own background") was used. As can be seen in Figure 27 and Figure 28, the background in Pythia is larger. A PEPSI challenge was then done in two ways. First, the multiplicities for positive and negative pions, kaons, and protons from the disNG production were unfolded with the Pmatrix from disNG. Here the charge separated matrices give perfect results since the unfolding was performed on the exact data that the ratios were taken from - this demonstrates that there are no bugs in the Pmatrix generating and unfolding codes (see Figures 49 to 57, pages 70 to 78). The likeness and charge combined Pmatrices are not perfect since you are unfolding, for example, positive pions with a Pmatrix element that is averaged over all charged pions. However, the results are still good. Secondly, the Pmatrix extracted from Pythia was used to unfold the disNG production. This is a sort of "simulation" of what happens in data; Real data (simulated by disNG), that is produced using its "own background", is unfolded with a MC (here Pythia) which was produced with its "own background". We use the difference in generators to simulate the difference between the MC generators and real physics. In this second case the PEPSI challenge comes out much worse, as we would expect. These differences demonstrate that, since we can not recreate exactly a "real physics" generator, we must use both disNG and Pythia (since we do not know which is closest to reality) Pmatrices and take the difference as a systematic error.

6.6 Geometry studies: 99 vs. 06

Productions were made using the 99 and the 06 geometry. The background levels in both disNG and Pythia are significantly higher in the 06 geometry then their 99 counterparts (see Figures 27 (page 55) and 28). This is verified by the difference seen in the overall background level of the different data taking periods (see Figure 25). Consequently, in both disNG and Pythia a significant difference in the Pmatrices is seen between 99 and 06 geometry.

6.7 Lambda decay studies

Extraction of proton Pmatrices for 1, 2, and 3+ tracks from lambda decays in the 00d2 data set were made by Yuri Naryshkine. The lambda peak was fit (with no RICH identification), a background subtraction made, and then the number of tracks in the peak was estimated. Then the procedure was repeated with the additional requirement that the proton track in the peak was identified as a pion (kaon, proton). The ratio of these two numbers was then made to construct the Pmatrices. A systematic error was estimated by varying the background and Lambda peak fitting ranges. This procedure was done for both IRT and EVT. In Figure 47 the IRT result is compared to various MC Pmatrices. In Figure 48 the EVT result is compared to various MC Pmatrices. The error bars represent the statistical error, the yellow bands represent the systematic error on the Lambda Pmatrices. It is important to note that the 1-track events are taken primarily from events with a short pion track in the same detector half, which may have some influence at the RICH. Also, the Lambda are extracted from the photoproduction region and have a very specific, non-standard topology. And additionally the results seem to be sensitive to the fit function used. Given these constraints, the lambda Pmatrices agree reasonably well with the MC Pmatrices, particularly for EVT.

7 Acknowledgments

Many many many thanks to Elke Aschenauer, Hal Jackson and Naomi Makins for long hours of discussion and work on this topic.

References

[1] E. Cisbani D. Schepper R. Kaiser. Particle Identification with the HERMES RICH Detector: Description of the Different Approaches. *HERMES internal note 98-008*, 1998.

[2] Jurgen Wendland. Particle identification for HERMES run i. *HERMES internal note 01-067*, 2001.

[3] A. Maas. RPS - RICH PID Scheduler. *HERMES internal note 00-029*, 2000.

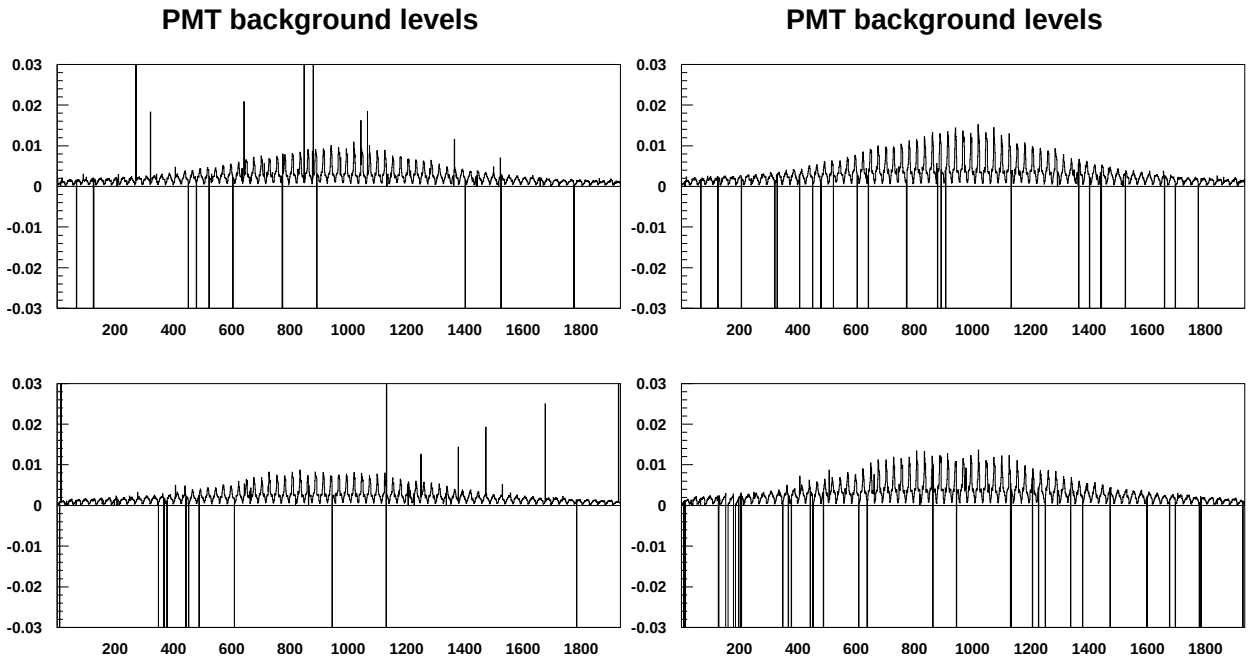

Figure 23: The background file for the top and bottom detectors as a function of the PMT number (1-1934). The negative values indicate a dead tube. The spikes indicate a hot tube and are set to a negative value before the file is used in a production.
(left) Extracted from the 2000 polarized data. (2000dataBKG)
(right) Extracted from 2007 data, all together

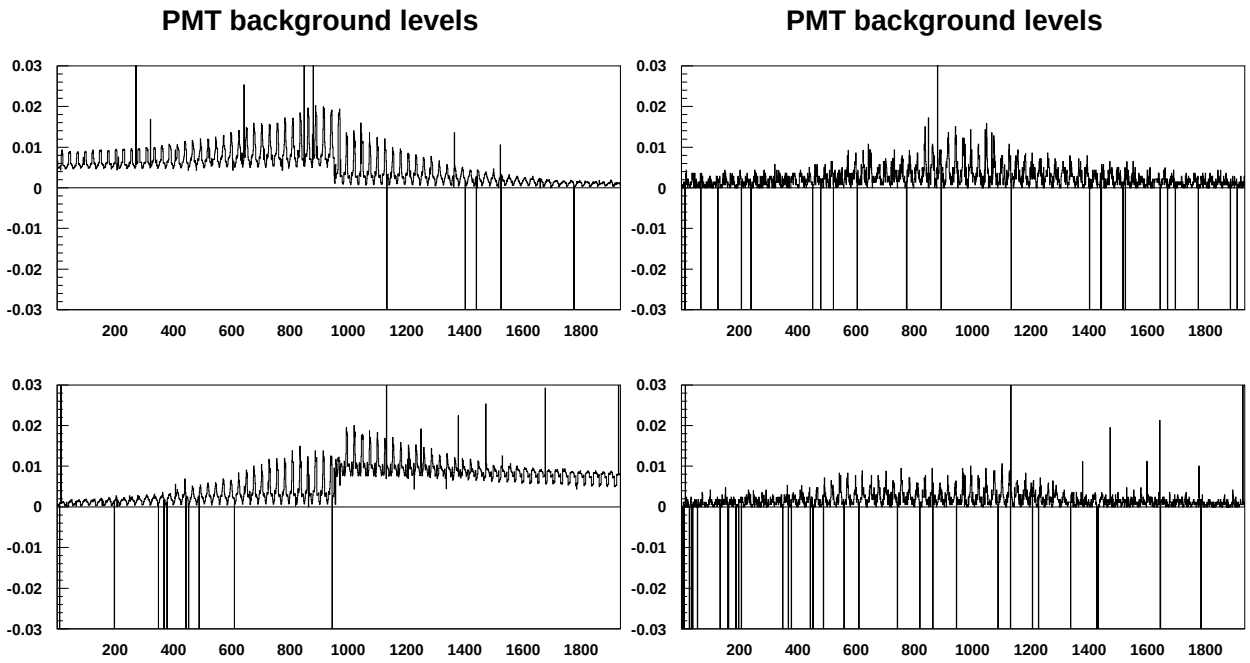

Figure 24: The background file for the top and bottom detectors as a function of the PMT number (1-1934). The negative values indicate a dead tube. The spikes indicate a hot tube.
(left) Extracted from the 2000 high density data. The step function around 1000 is discussed in the text.
(right) Extracted from the 1999 data. The "noisy" behavior is not understood and thus this file is not used for data productions.

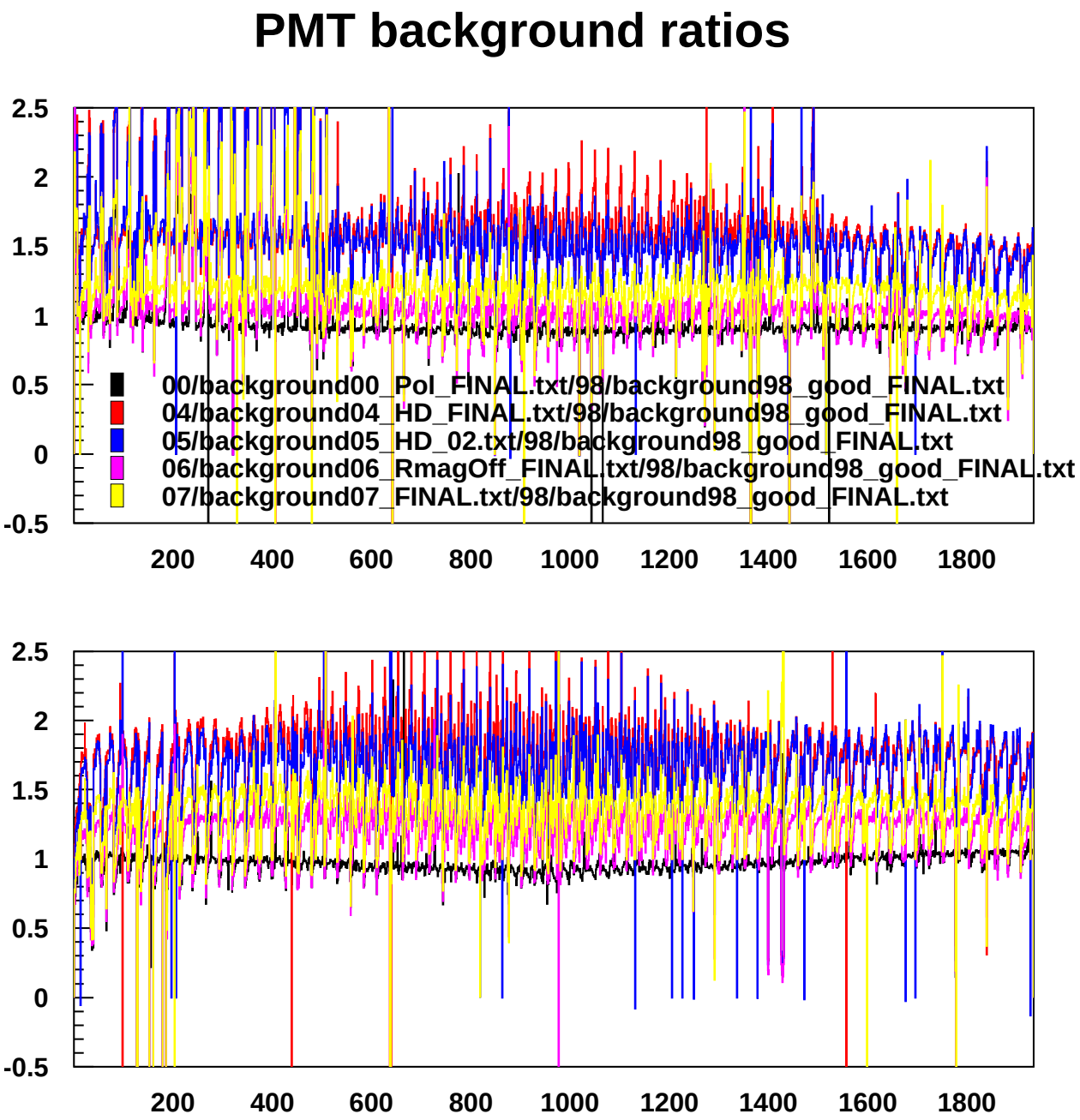

Figure 25: The ratio of background files, various years to the 98 background. The overall values seem to group by year: 98 and 00 have the lowest average, 04 and 05 have the largest, and 06 and 07 fall somewhere in the middle. However, it should be noted that 04 and 05 are taken from HD data while for all other years the background was taken from only polarized data or all data mixed together. See section 4.1.1.

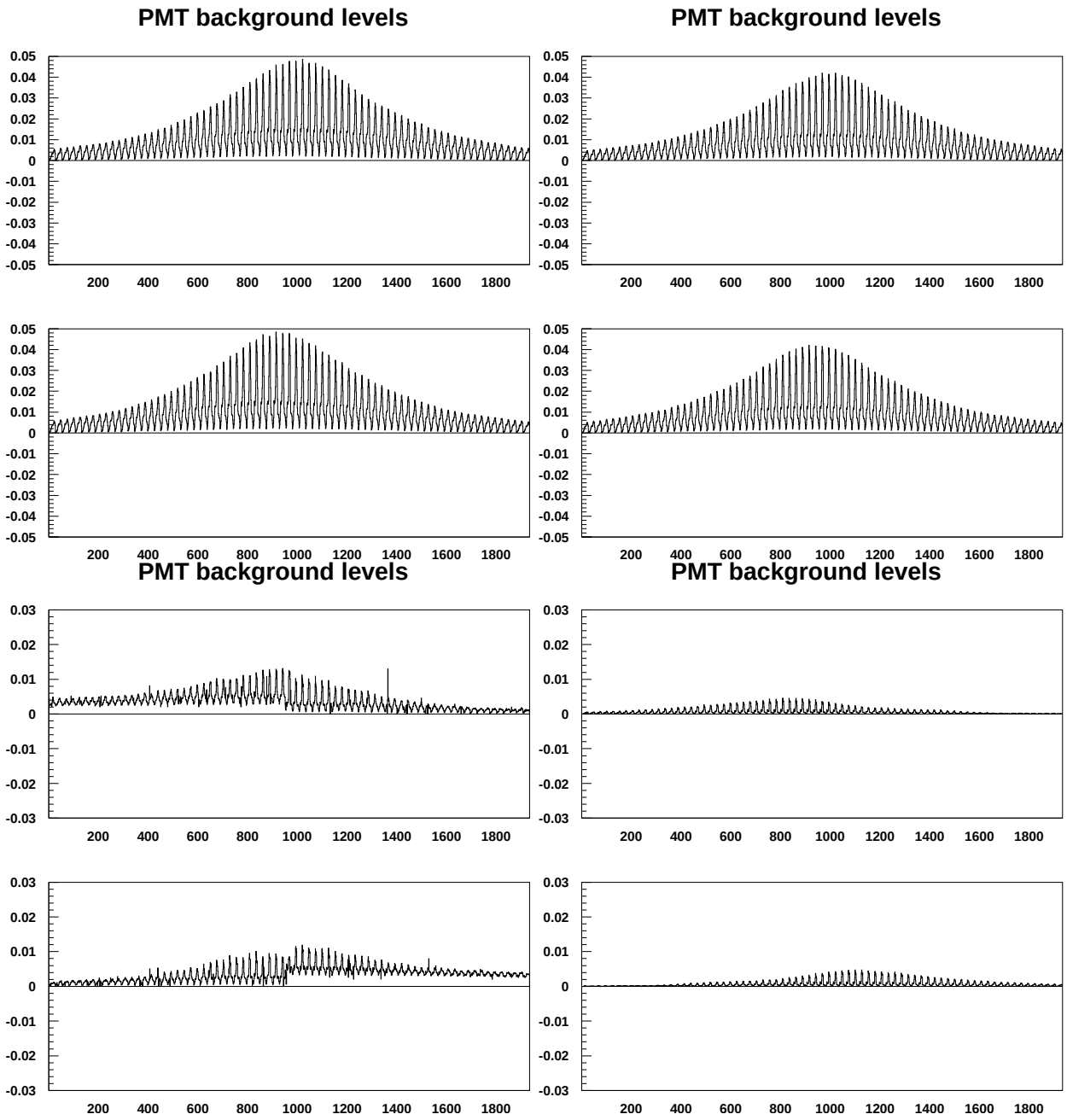

Figure 26: Background files used in Preliminary MC Studies (see Section 5)
(top left) The background files extracted from a pythia MC sample produced with a selector on K_s particles with dobeampipe ON and (top right) dobeampipe OFF.
(bottom left) The "smoothed" background file extracted from the 2000 data.
(bottom right) The background file extracted from a (mystery) disNG MC sample (HRC files taken from Tokyo)

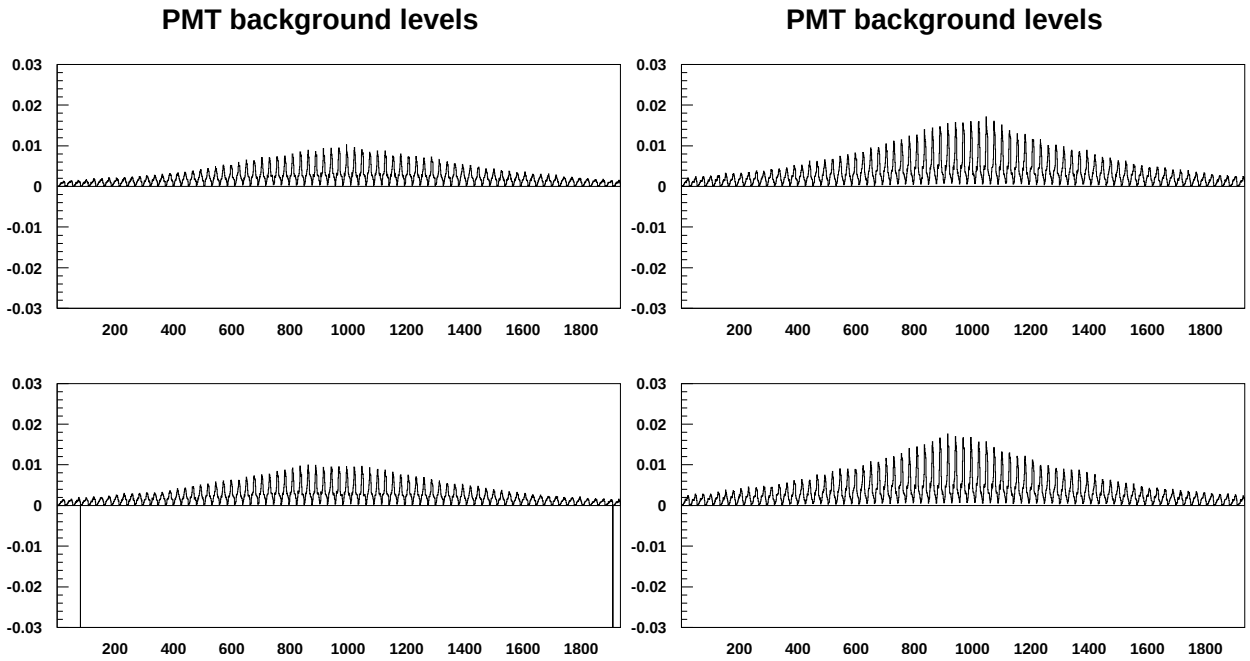

Figure 27: (left) The background file from a disNG production with 99 geometry (→ disNG99 BKG).

(right) The background file from a pythia production with 99 geometry (→ Pythia99 BKG).

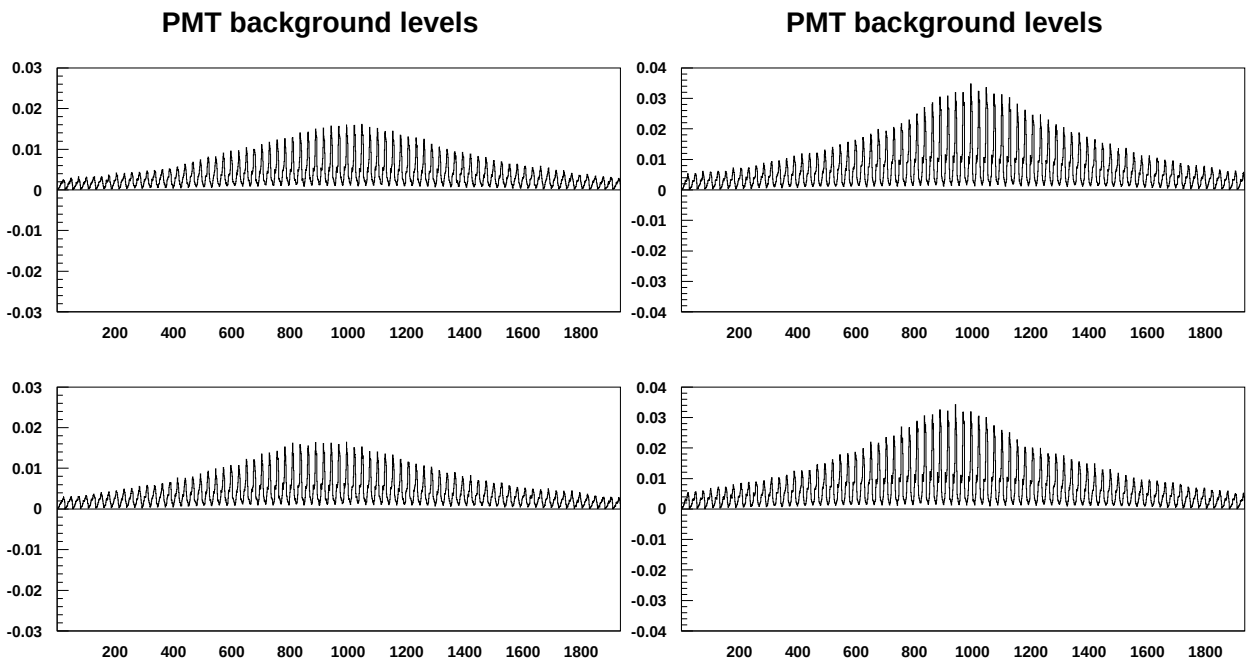

Figure 28: (left) The background file from a disNG production with 06 geometry (→ disNG06 BKG).
(right) The background file from a Pythia production with 06 geometry (→ Pythia06 BKG). Note larger scale.

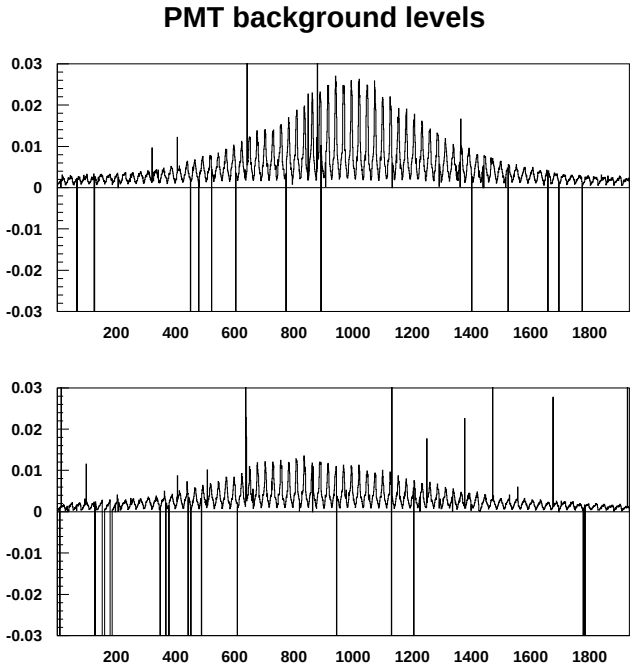

Figure 29: The background file from the 2005 Polarized data, used as a "worst case" in the background study (→ polData2005 BKG)

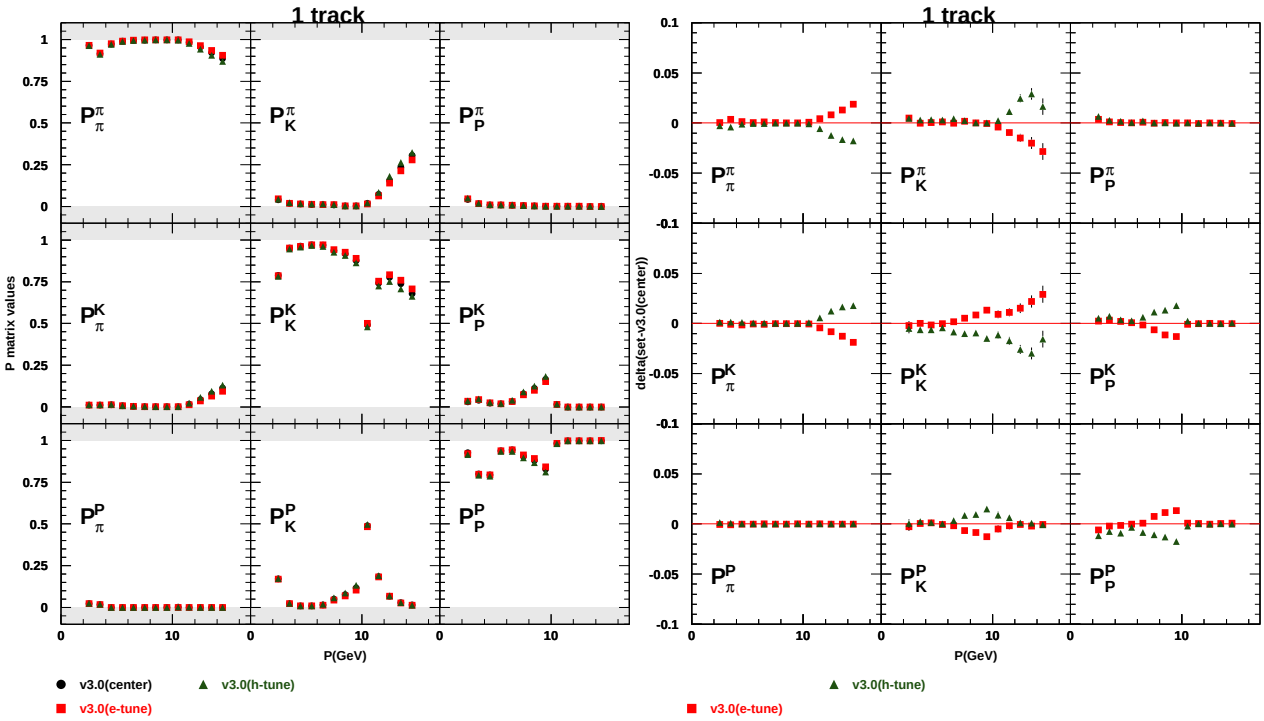

Figure 30: The 1 track version 3.0 Pmatrices (99 geometry, IRT, charge combined) for the e-, h-, and c-tunes and (right) the difference of h- and e-tunes to the c-tune

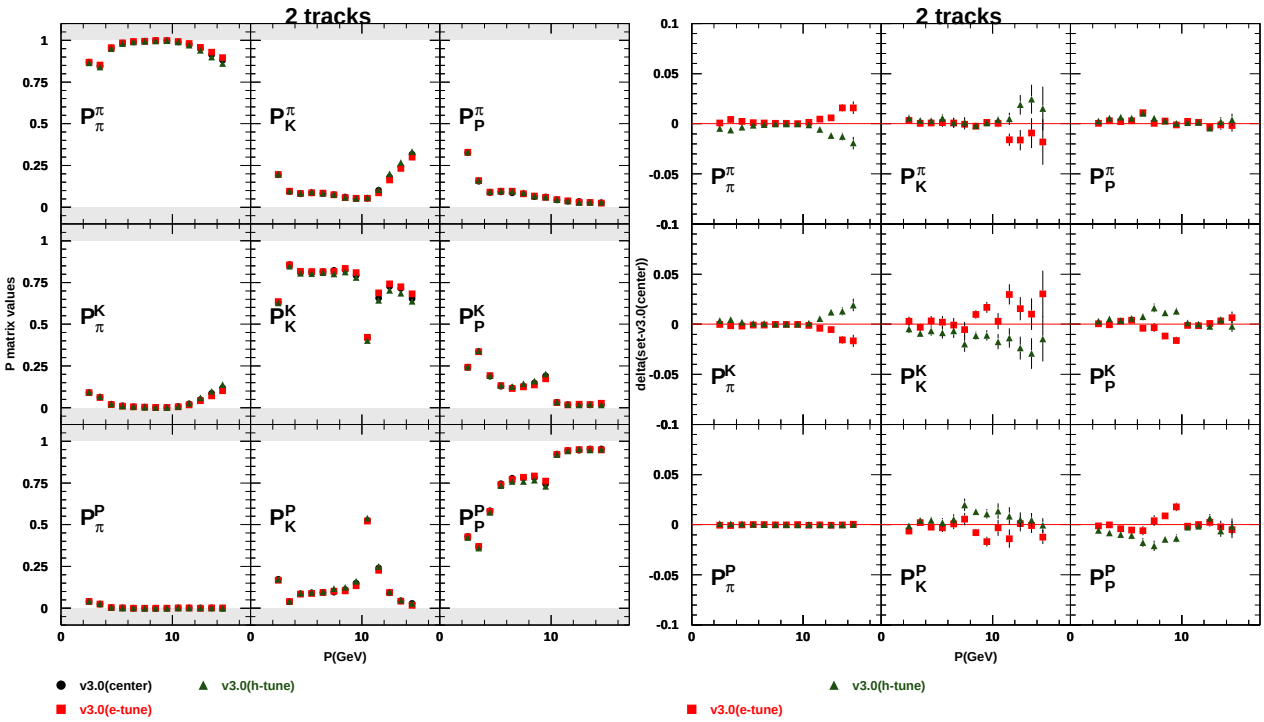

Figure 31: The 2 track version 3.0 Pmatrices for the e-, h-, and c-tunes and (right) the difference of h- and e-tunes to the c-tune

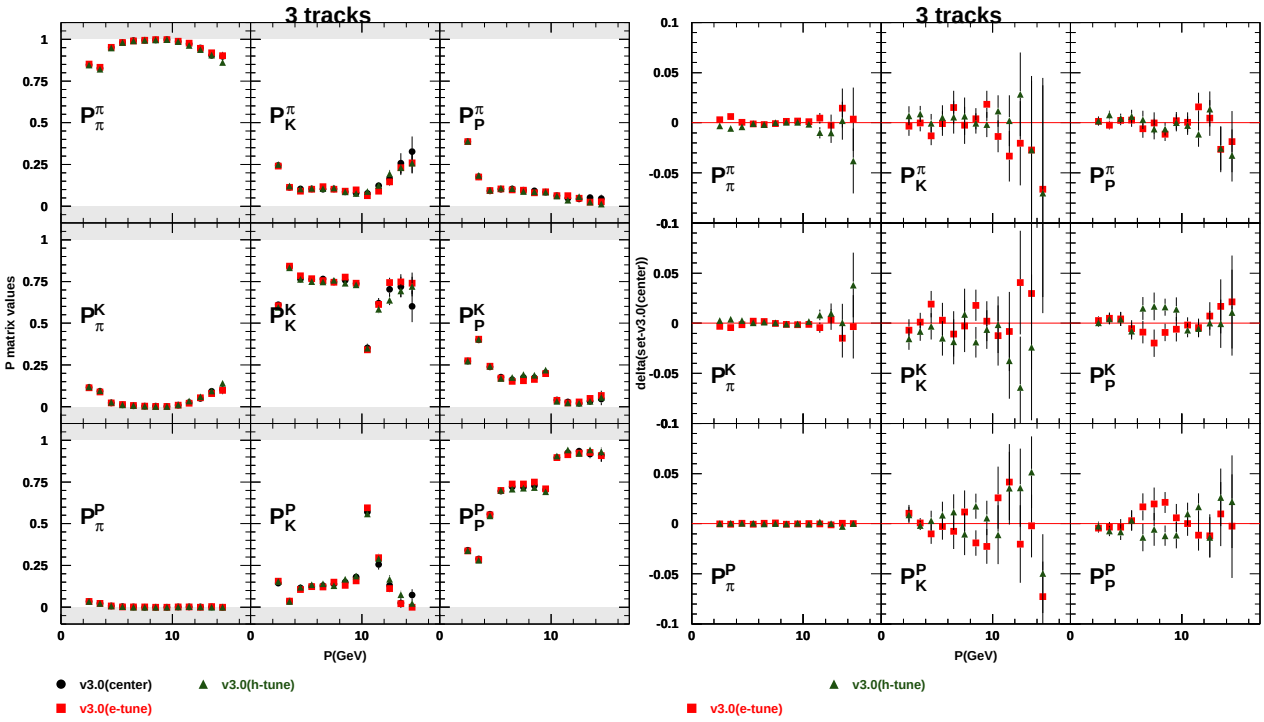

Figure 32: The 3+ track version 3.0 Pmatrices for the e-, h-, and c-tunes and (right) the diference of h- and e-tunes to the c-tune

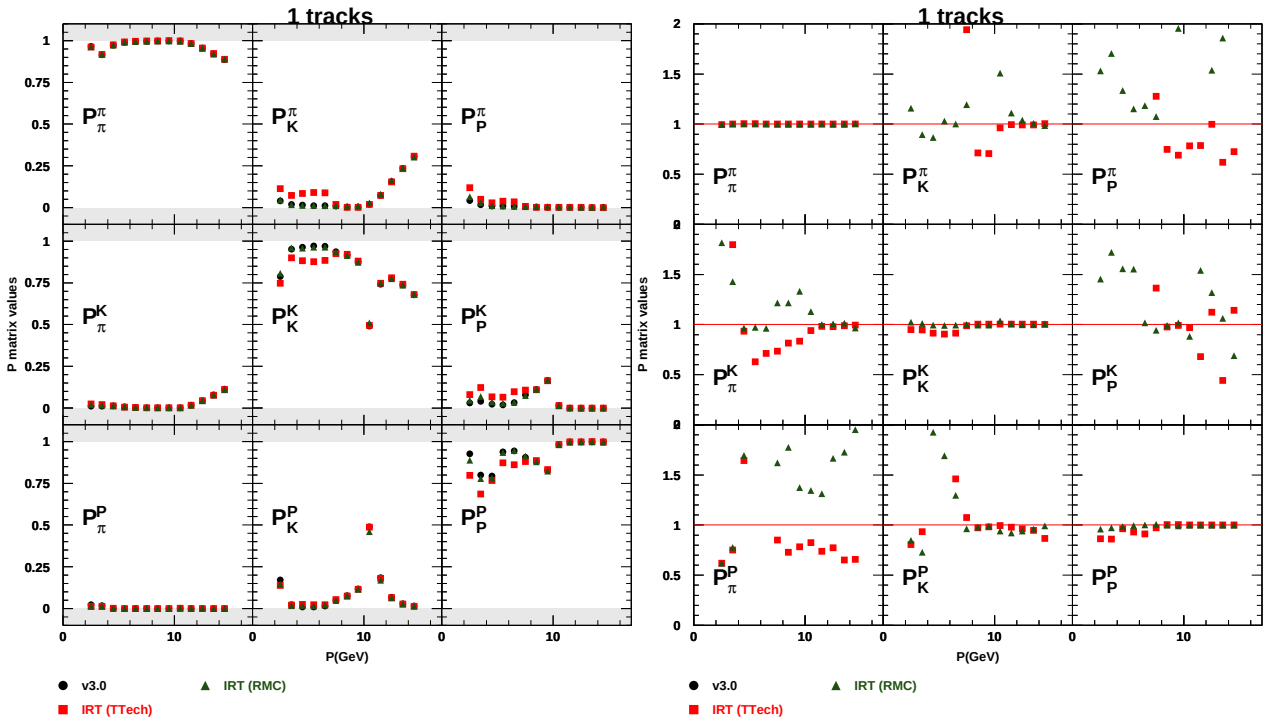

Figure 33: Comparison of various charge combined ("all") IRT Pmatrices (left) for the case of 1 track in the respective detector half. Shown are the v3.0 IRT Pmatrix, a new Pmatrix extraction from the same Monte Carlo production using the current cuts ("IRT (TTech)") and Pmatrices extracted from a new disNG 1999 geometry MC production run with disNG99 background and using with selector files. The ratios of the latter two Pmatrices to v3.0 are shown on the right.

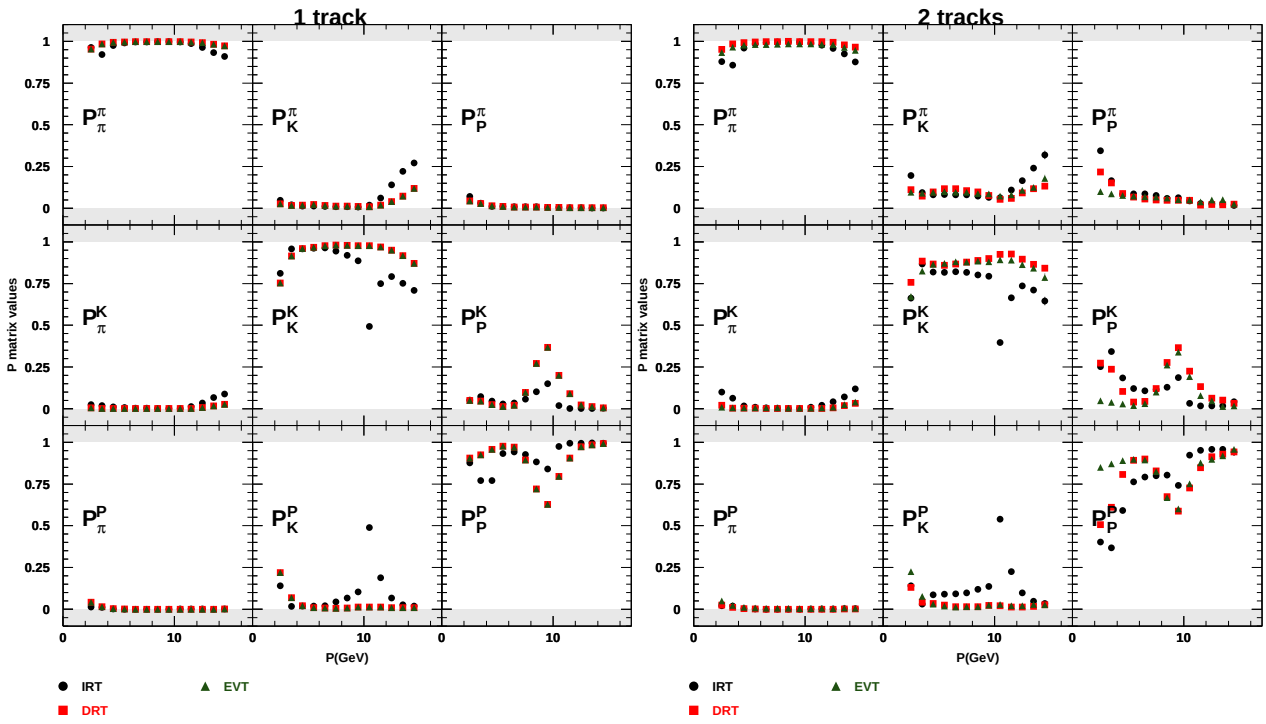

Figure 34: Comparison of the IRT, DRT, and (for 2 track) EVT charge combined ("all") Pmatrices extracted from the disNG 2006 geometry production run with the disNG06 background for one (left) and two (right) tracks in the detector half (for 1 track, EVT=DRT).

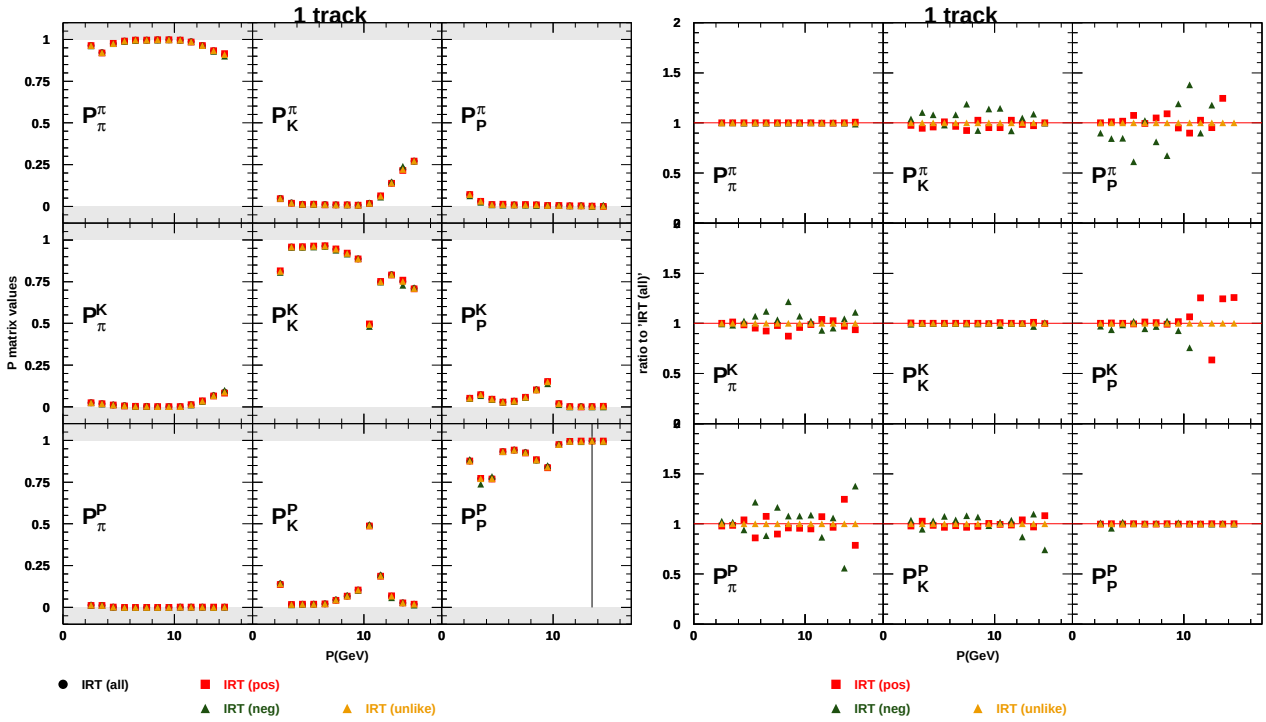

Figure 35: Comparison of the IRT charge combined (all), charge separated (positive and negative) and likeness separated (like and unlike) Pmatrices extracted from the disNG 2006 geometry production.

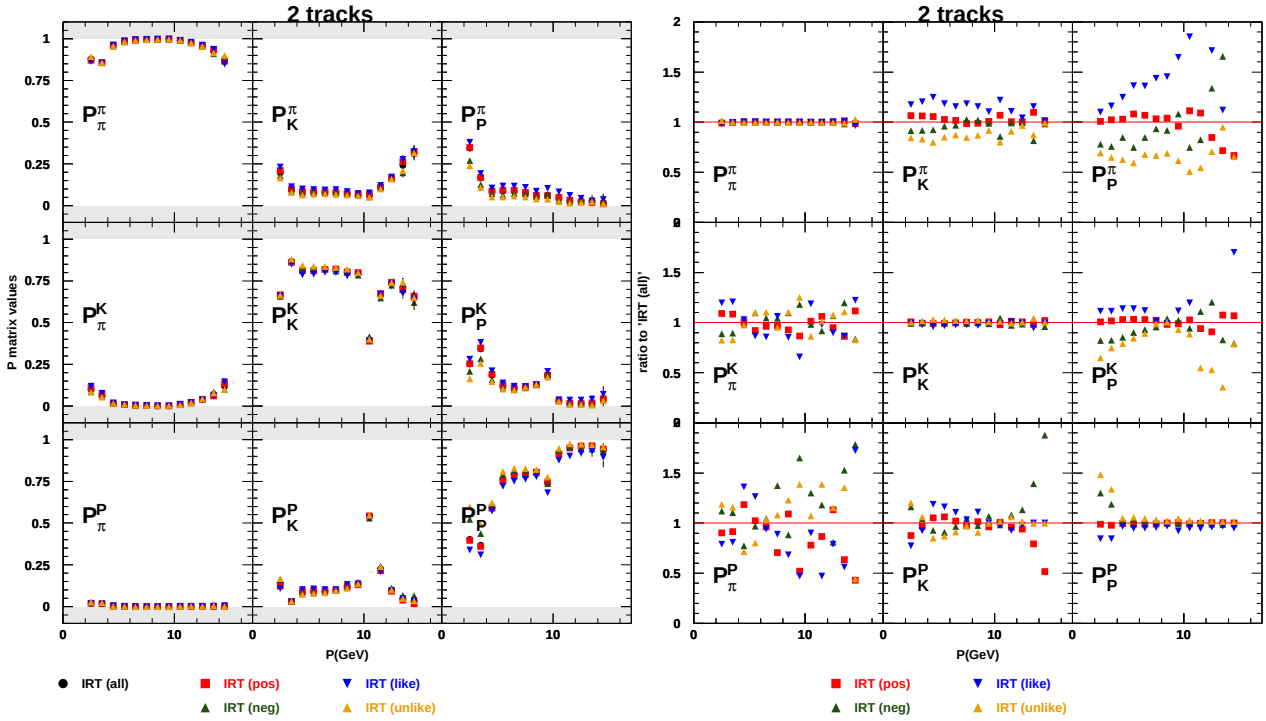

Figure 36: Same as Figure 35, but for two tracks in the detector half.

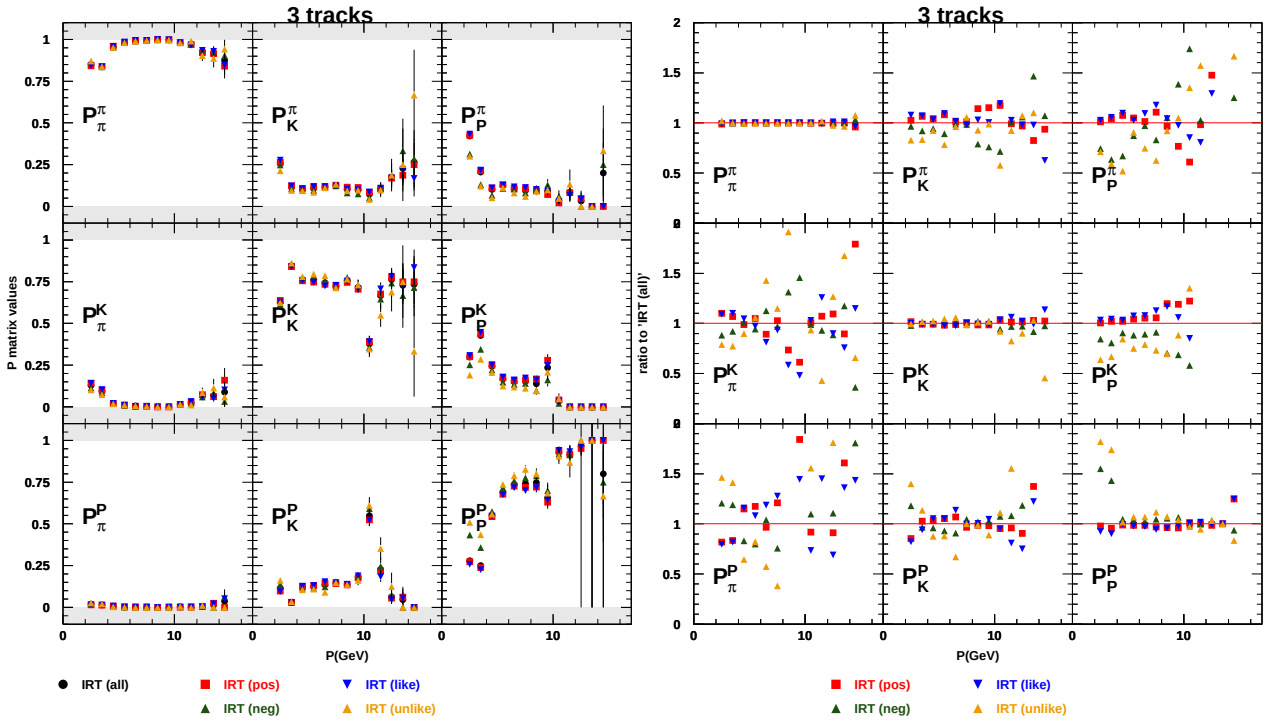

Figure 37: Same as Figure 35, but for three tracks in the detector half.

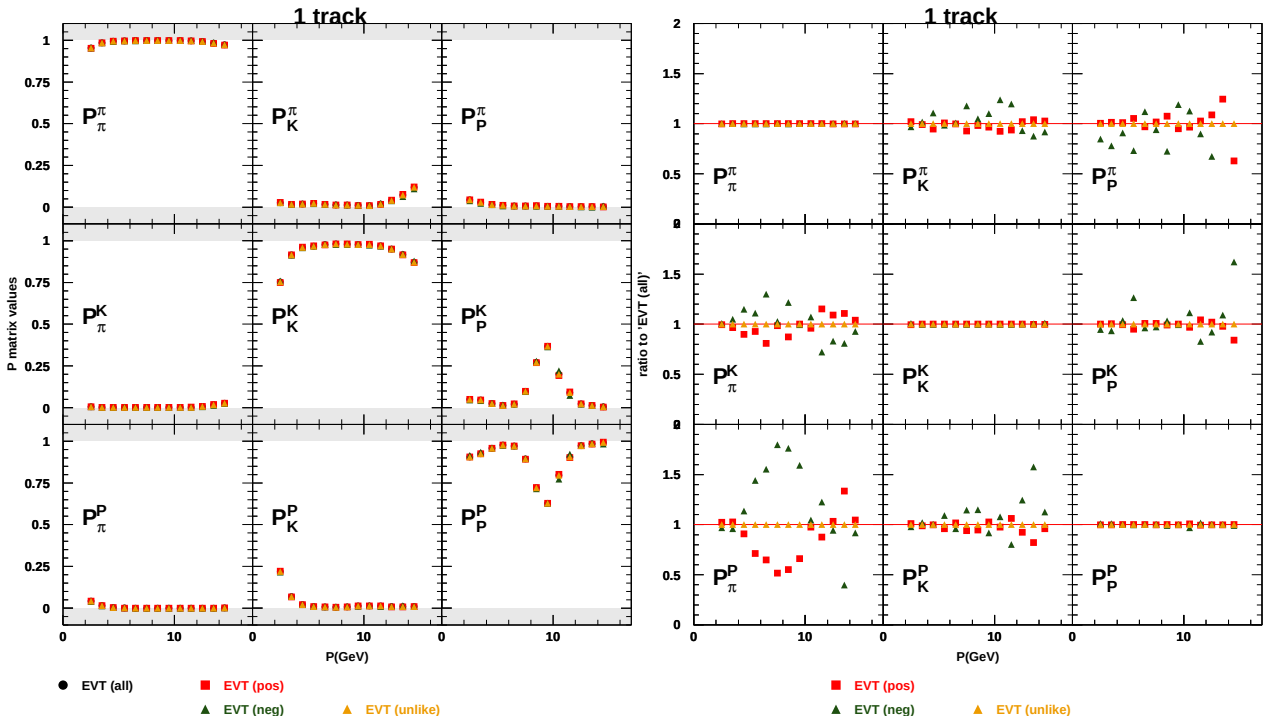

Figure 38: Comparison of the EVT charge combined (all), charge separated (positive and negative) and likeness separated (like and unlike) Pmatrices extracted from the disNG 2006 geometry production produced using the disNG06 background file.

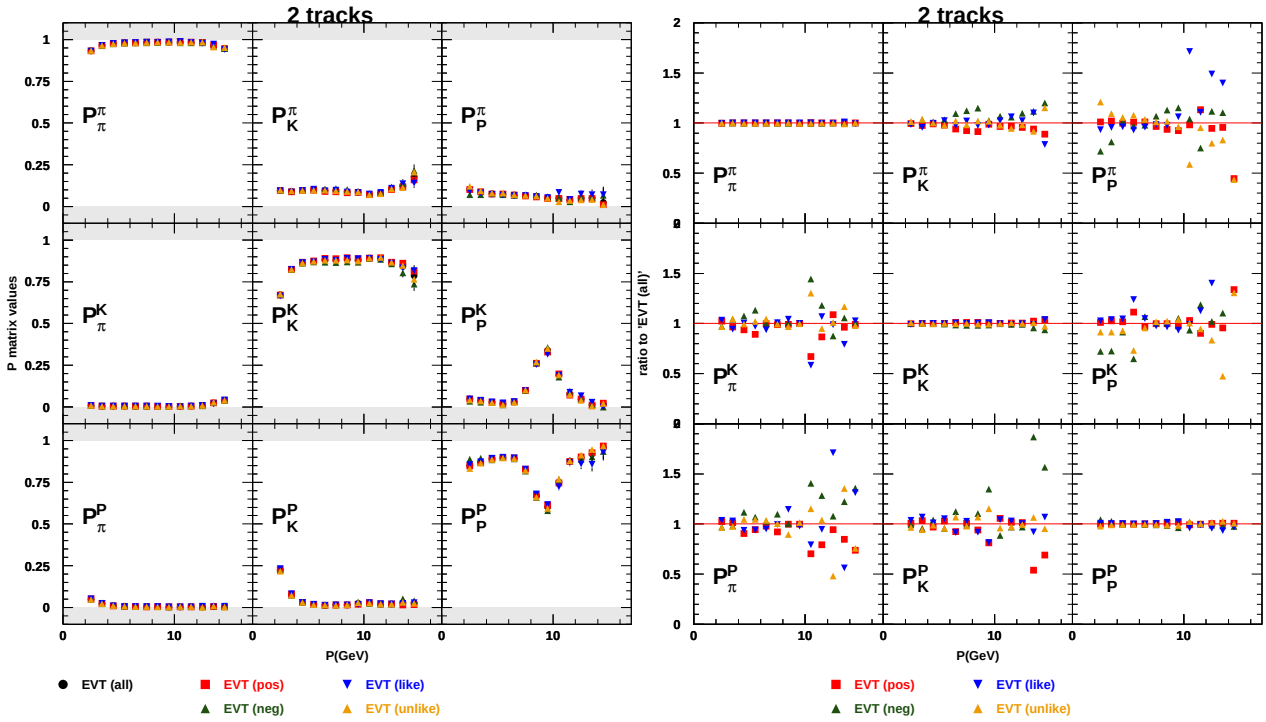

Figure 39: Same as Figure 38, but for two tracks in the detector half.

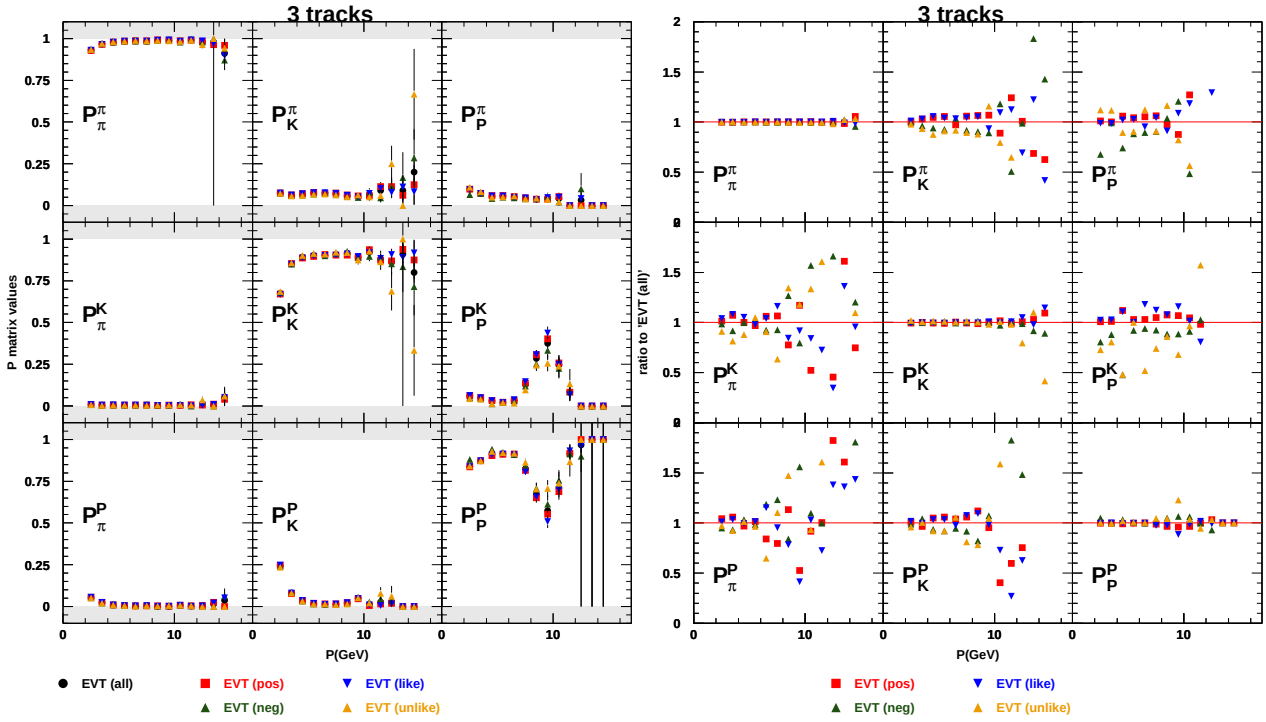

Figure 40: Same as Figure 38, but for three tracks in the detector half.

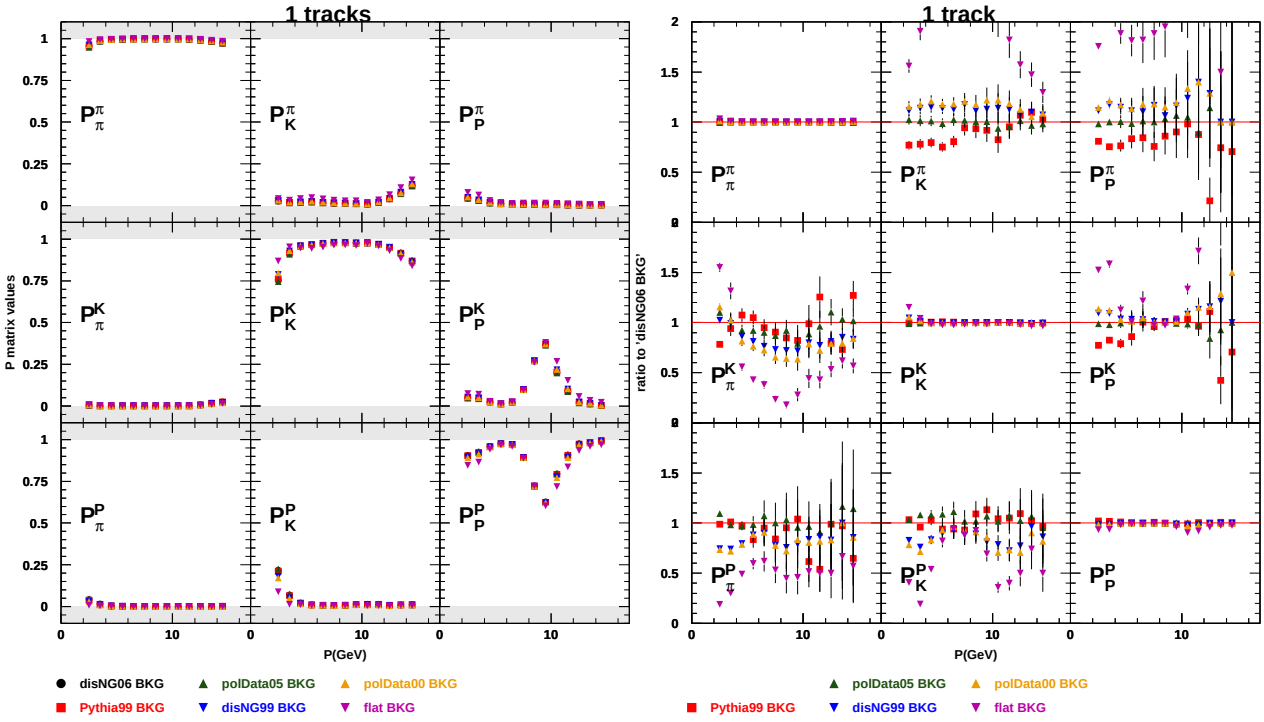

Figure 41: Charge combined ("all") Pmatrices and ratios of Pmatrices (1 track) for a disNG production with 2006 geometry, obtained for the EVT method using different background assumptions (see section 6.4). The ratios are relative to using the background extracted from the MC production itself (DISNG06 BKG). Lower background assumptions reduce the pion misidentifications as kaons and protons, but increase the misidentifications of heavier hadrons as pions.

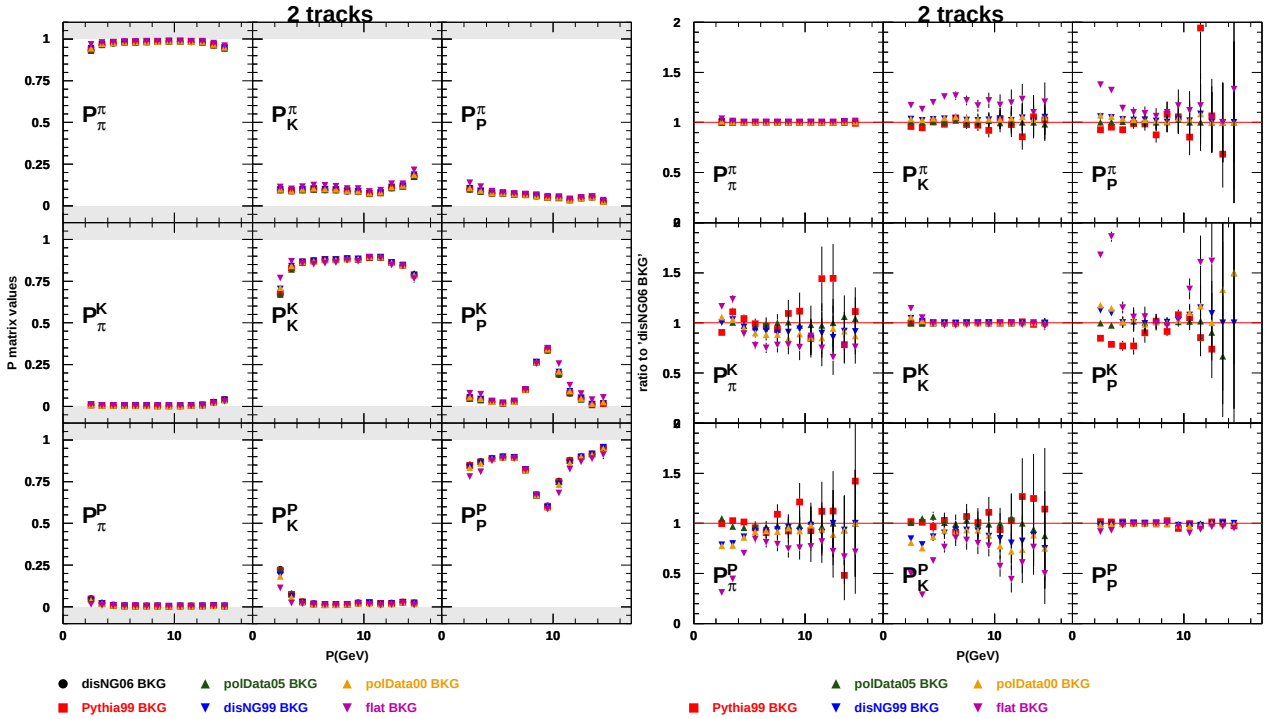

Figure 42: Same as Figure 41, but for two tracks in the detector half. Similar effects can be observed, but they are less pronounced.

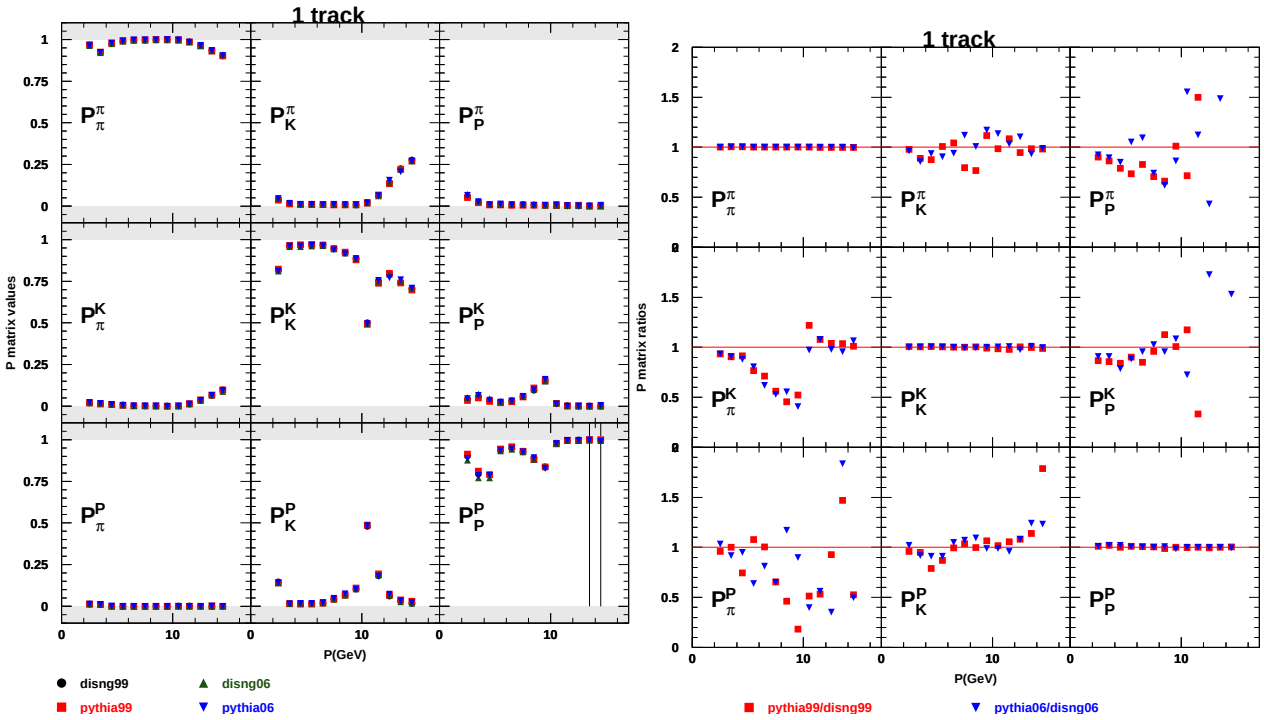

Figure 43: Charge combined (all) IRT Pmatrices and ratios of Pmatrices (1 track) for different generators and geometries. All production were produced with their own respective background files.

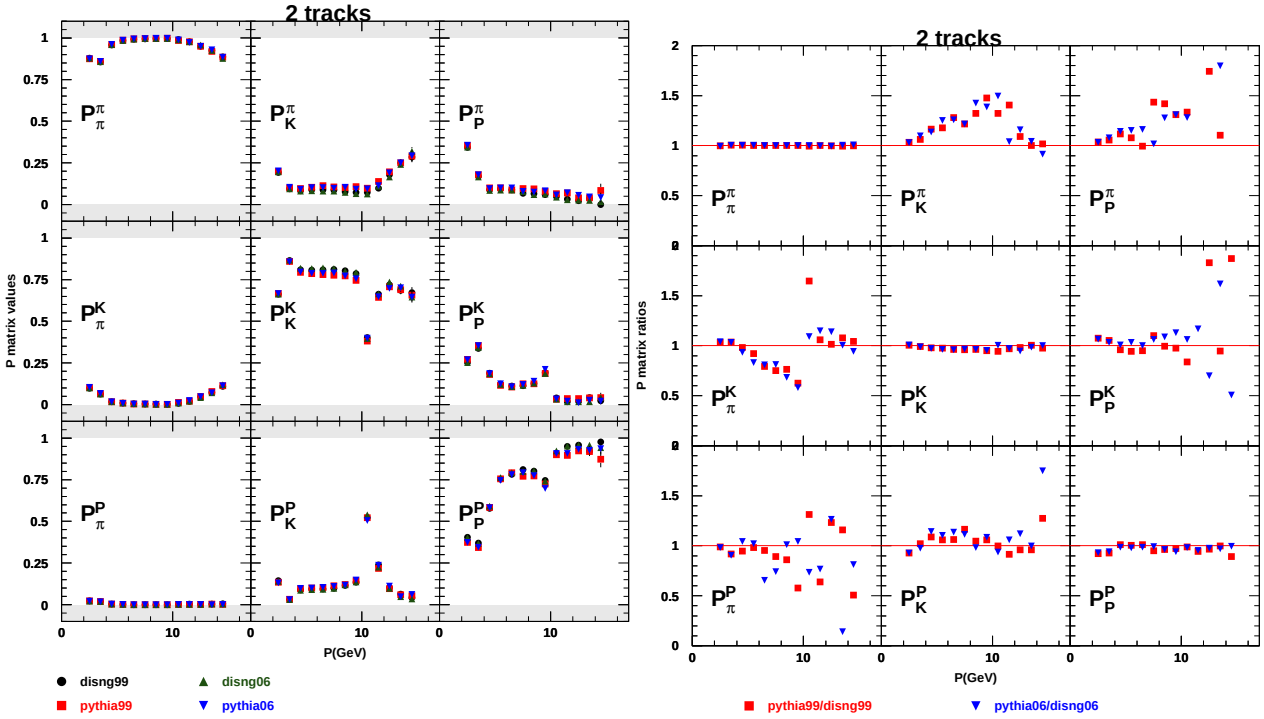

Figure 44: Same as Figure 43, but for two tracks in the detector half.

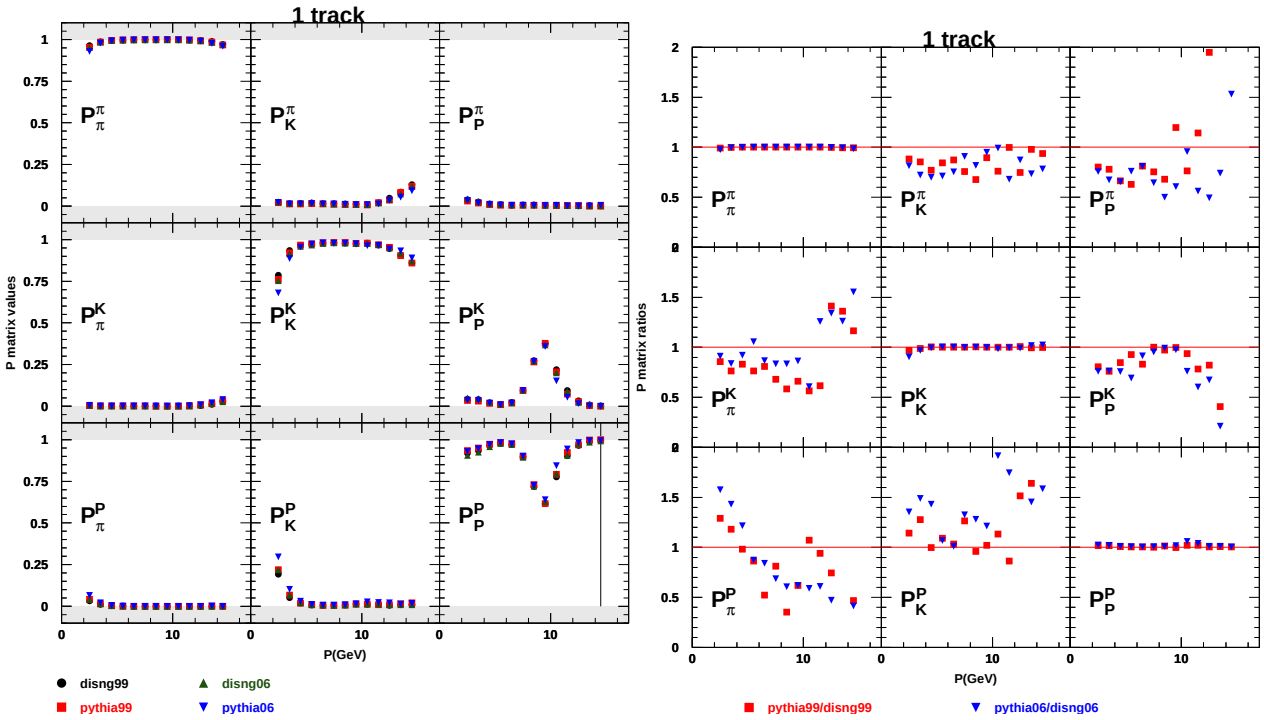

Figure 45: Charge mixed (all) EVT Pmatrices and ratios of Pmatrices (1 track) for a different generators and geometries. All production were produced with their own respective background files.

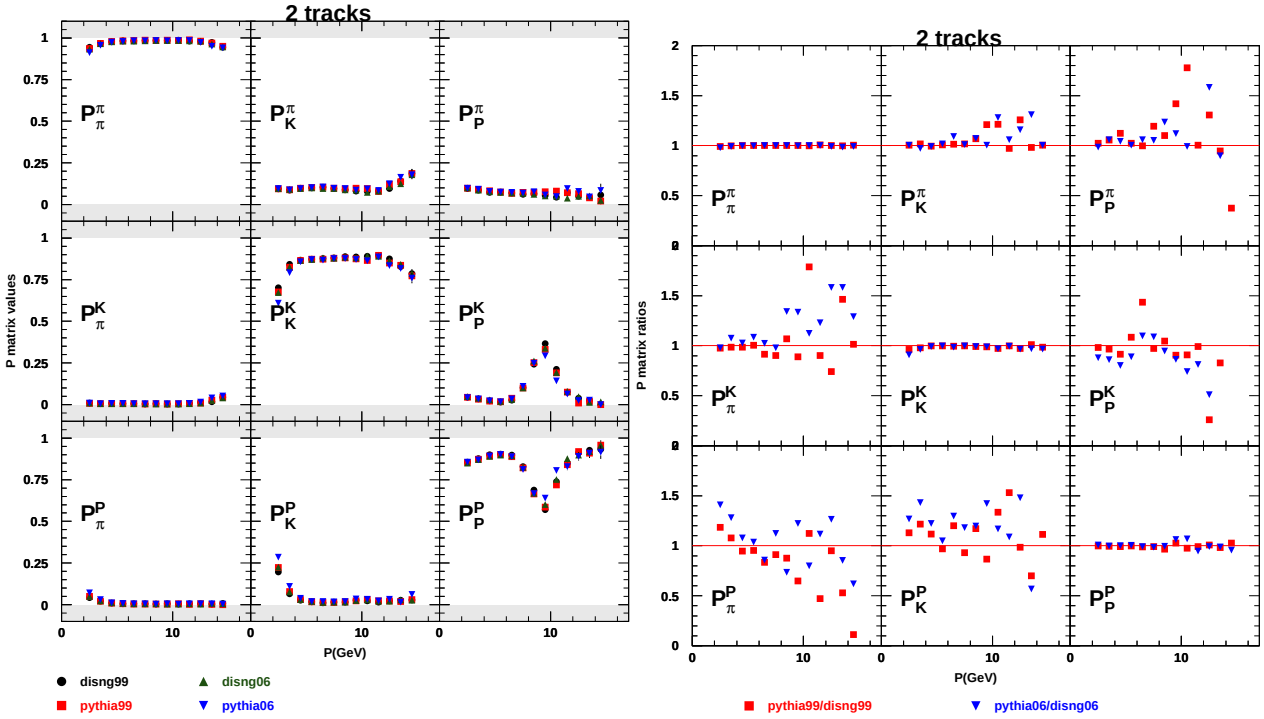

Figure 46: Same as Figure 45, but for two tracks in the detector half.

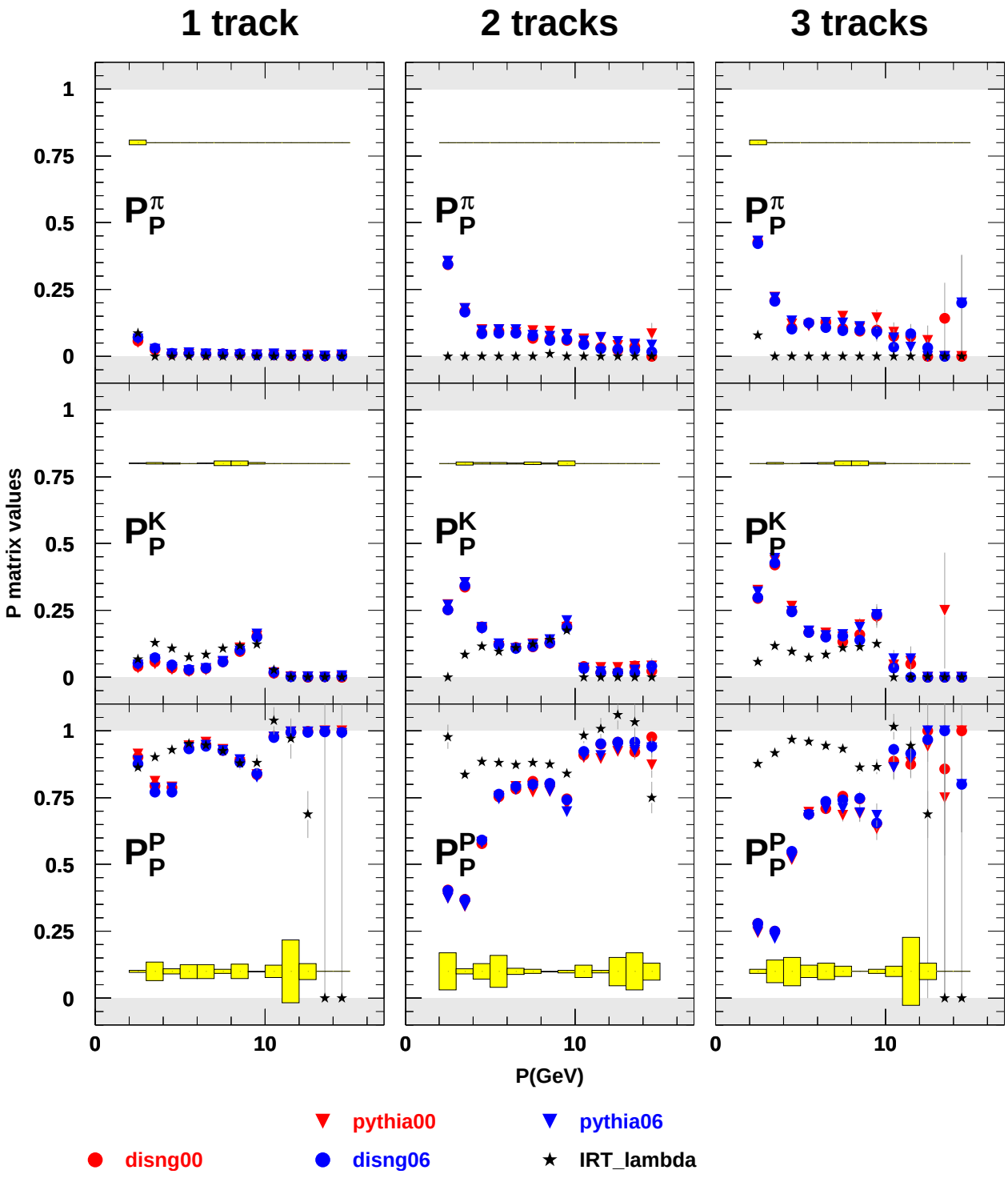

Figure 47: The proton columns of the charge combined (all) IRT Pmatrices plotted for Lambdas extracted from the 2000 data compared to MC Pmatrices from various geometries and generators. The yellow band gives the systematic error on the Lambda Pmatrices, estimated by varying the background and Lambda peak fitting ranges.

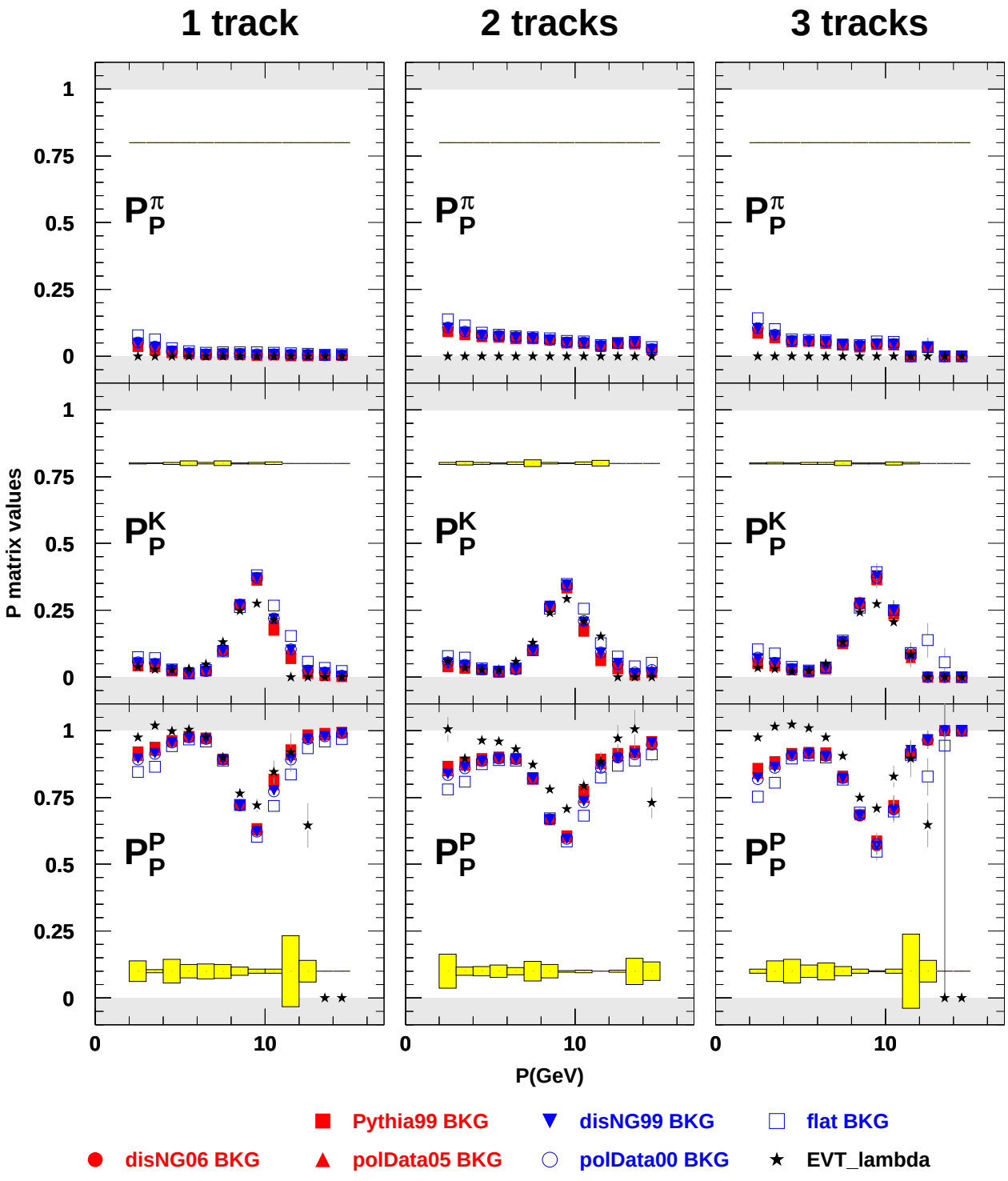

Figure 48: The proton columns of the charge combined (all) EVT Pmatrices plotted for Lambdas extracted from the 2000 data compared to MC Pmatrices from disNG06 with various background files (see section 6.4). The yellow band gives the systematic error on the Lambda Pmatrices, estimated by varying the background and Lambda peak fitting ranges.

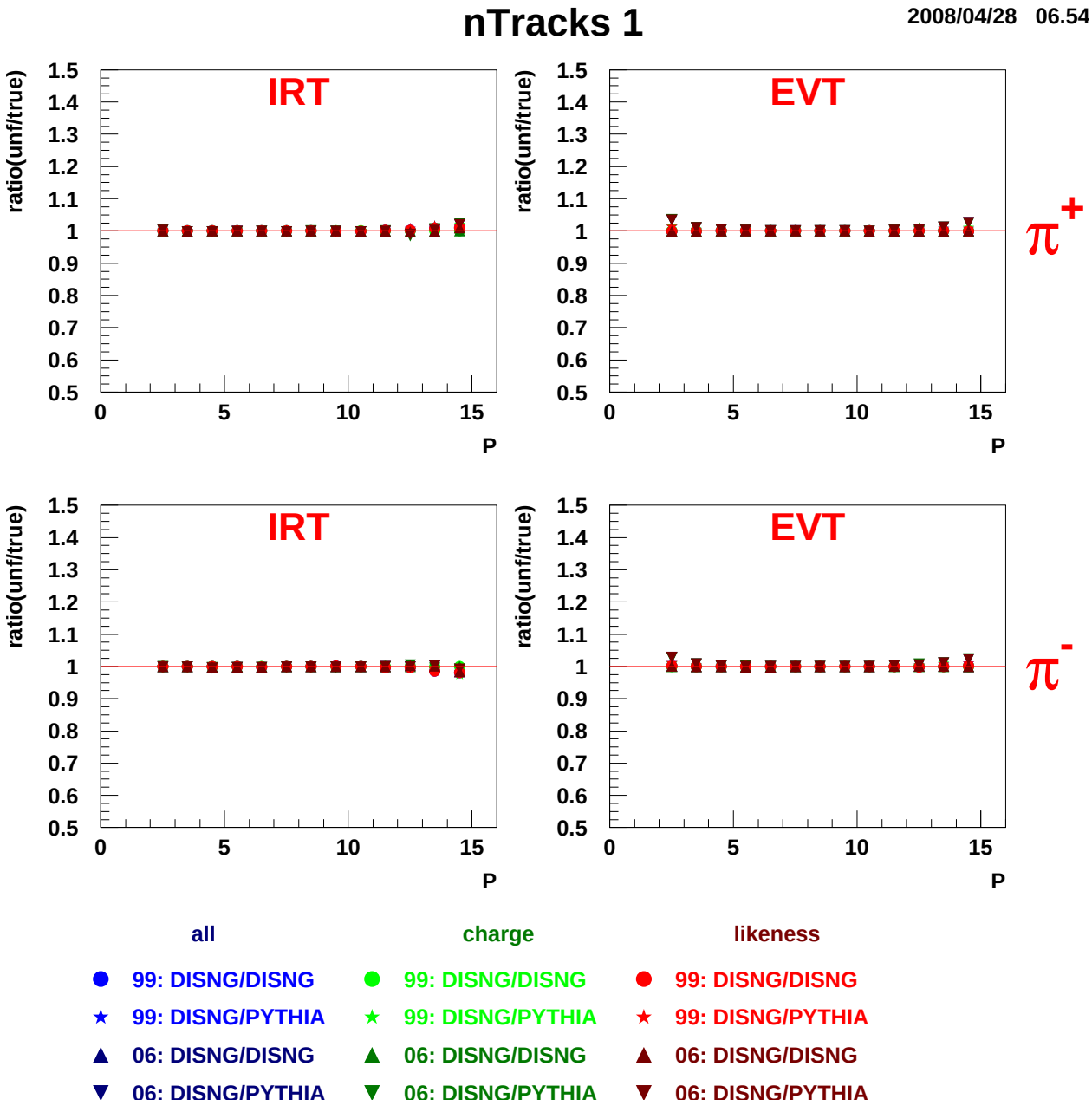

Figure 49: PEPSI Challenge for IRT and EVT (DRT) for 1 track in the detector half for positive and negative pions. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

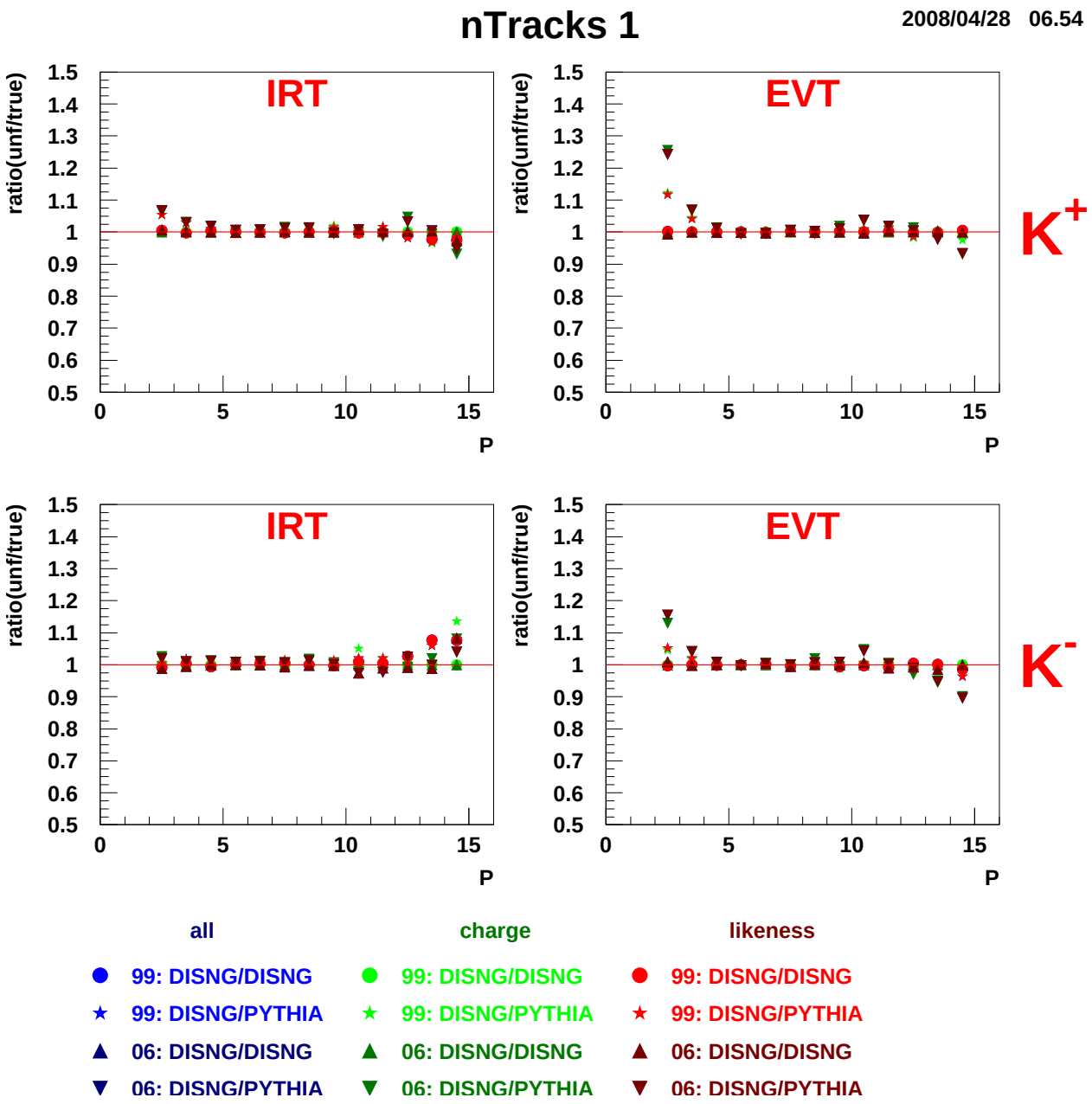

Figure 50: PEPSI Challenge for IRT and EVT (DRT) for 1 track in the detector half for positive and negative kaons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

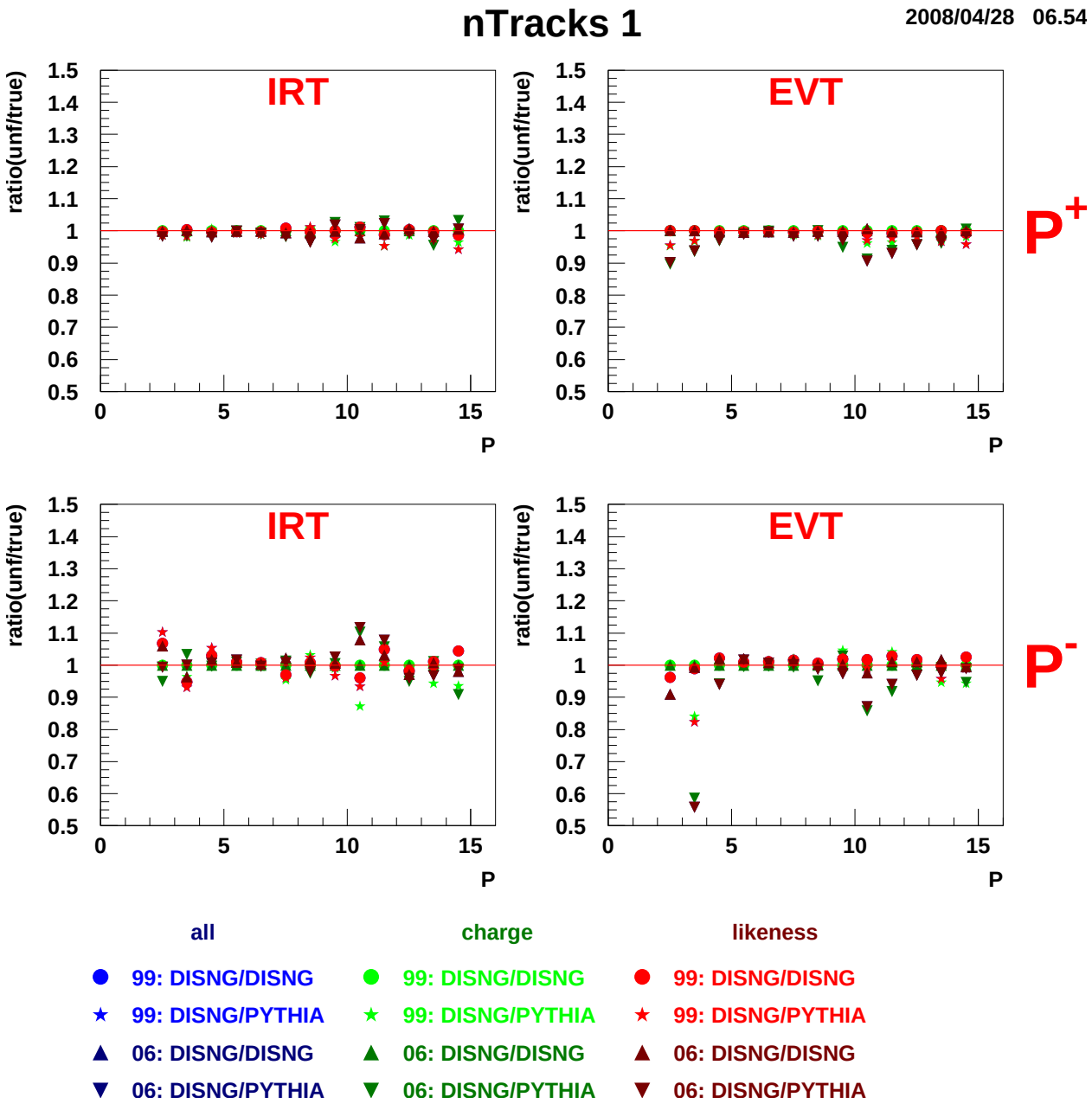

Figure 51: PEPSI Challenge for IRT and EVT (DRT) for 1 track in the detector half for protons and antiprotons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

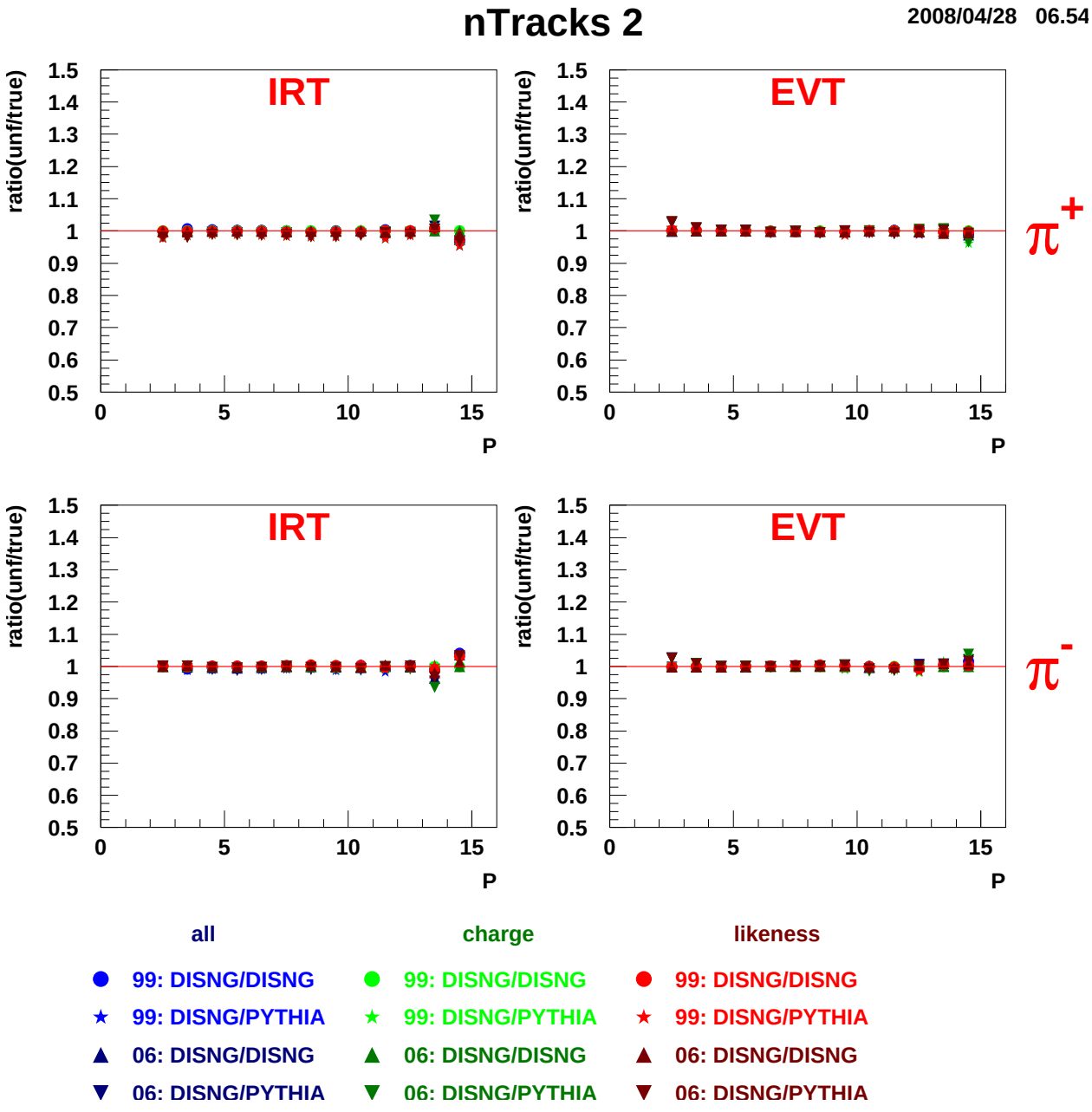

Figure 52: PEPSI Challenge for IRT and EVT for 2 tracks in the detector half for positive and negative pions. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

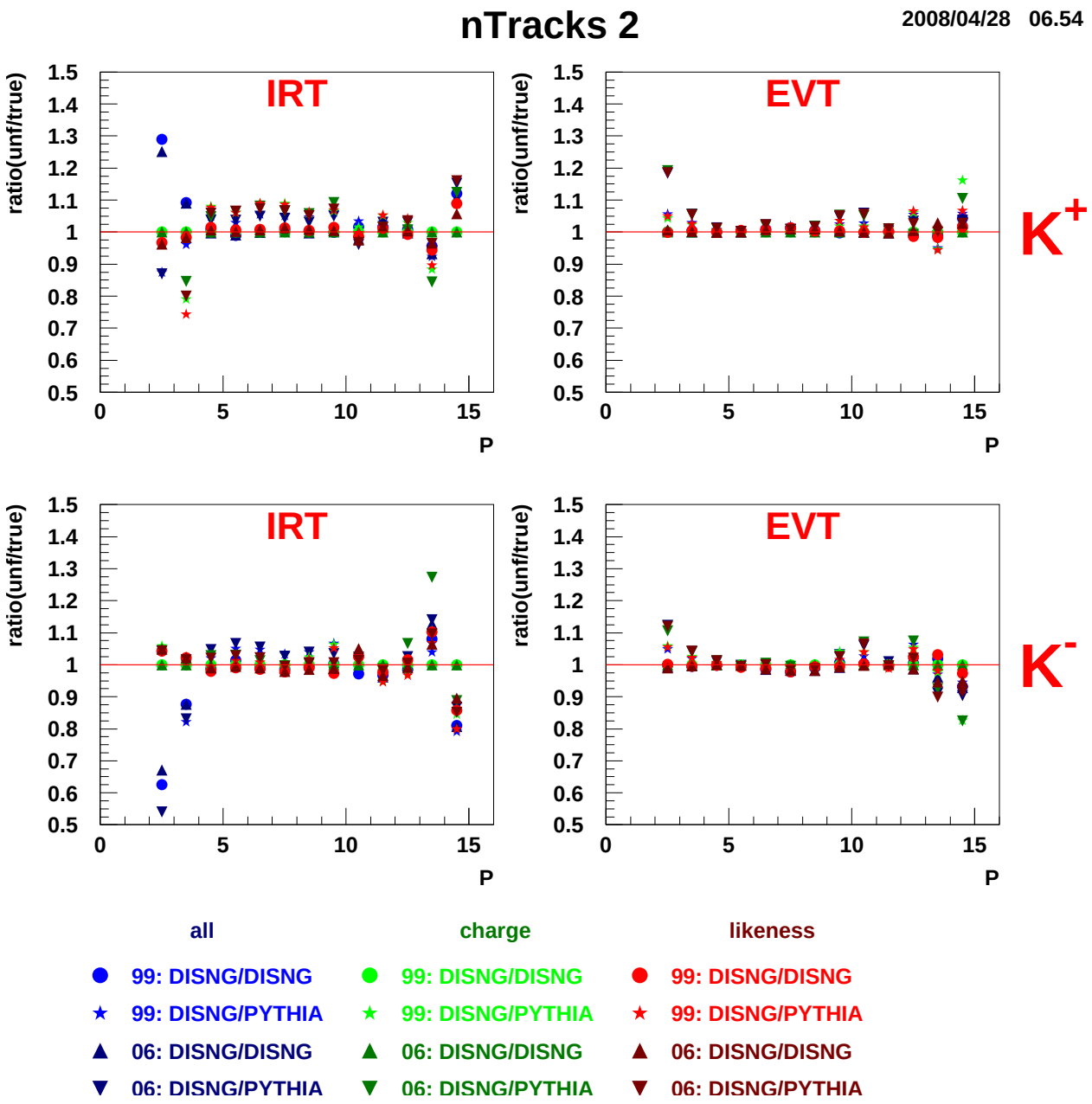

Figure 53: PEPSI Challenge for IRT and EVT for 2 tracks in the detector half for positive and negative kaons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

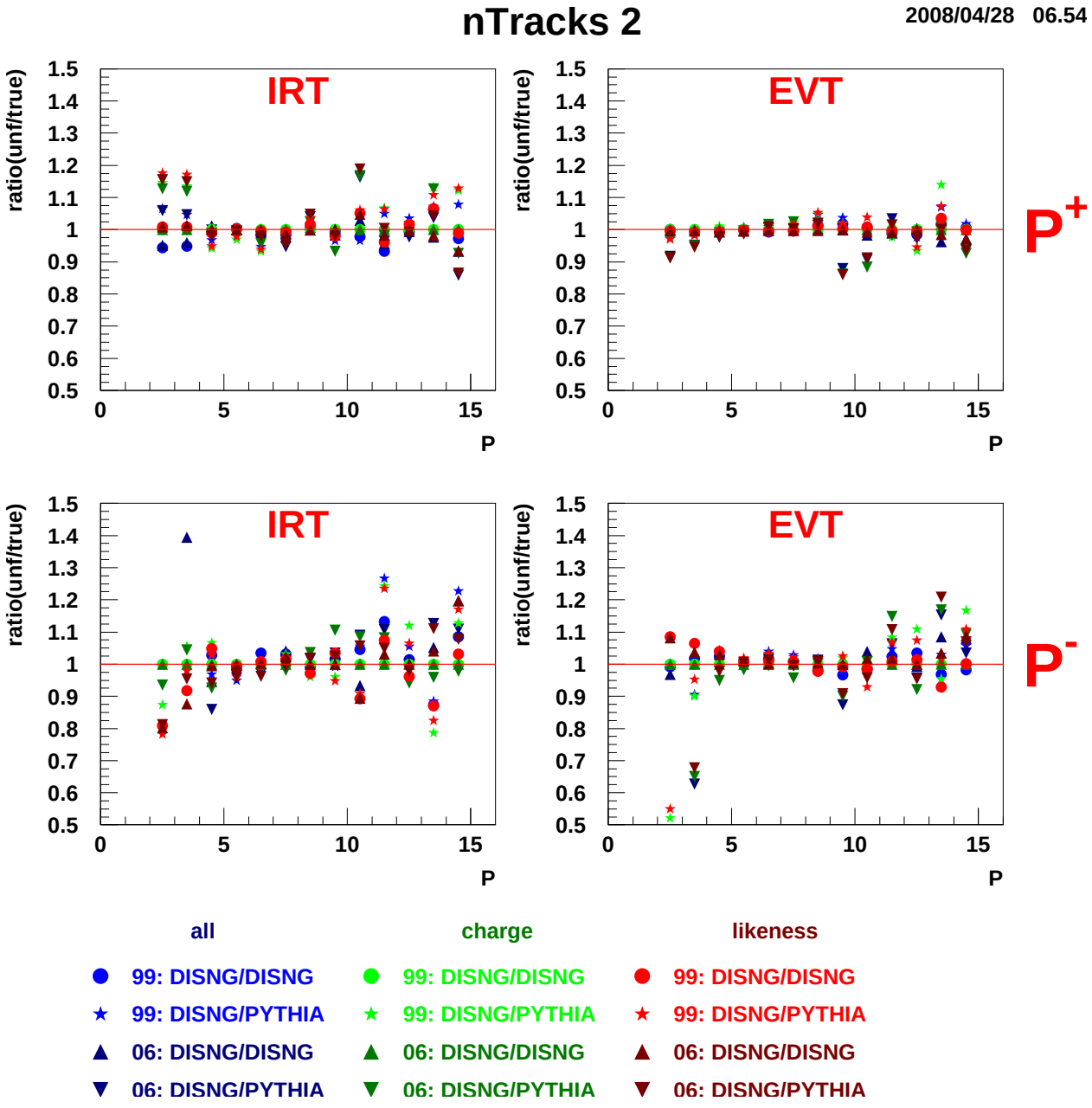

Figure 54: PEPSI Challenge for IRT and EVT for 2 tracks in the detector half for protons and antiprotons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

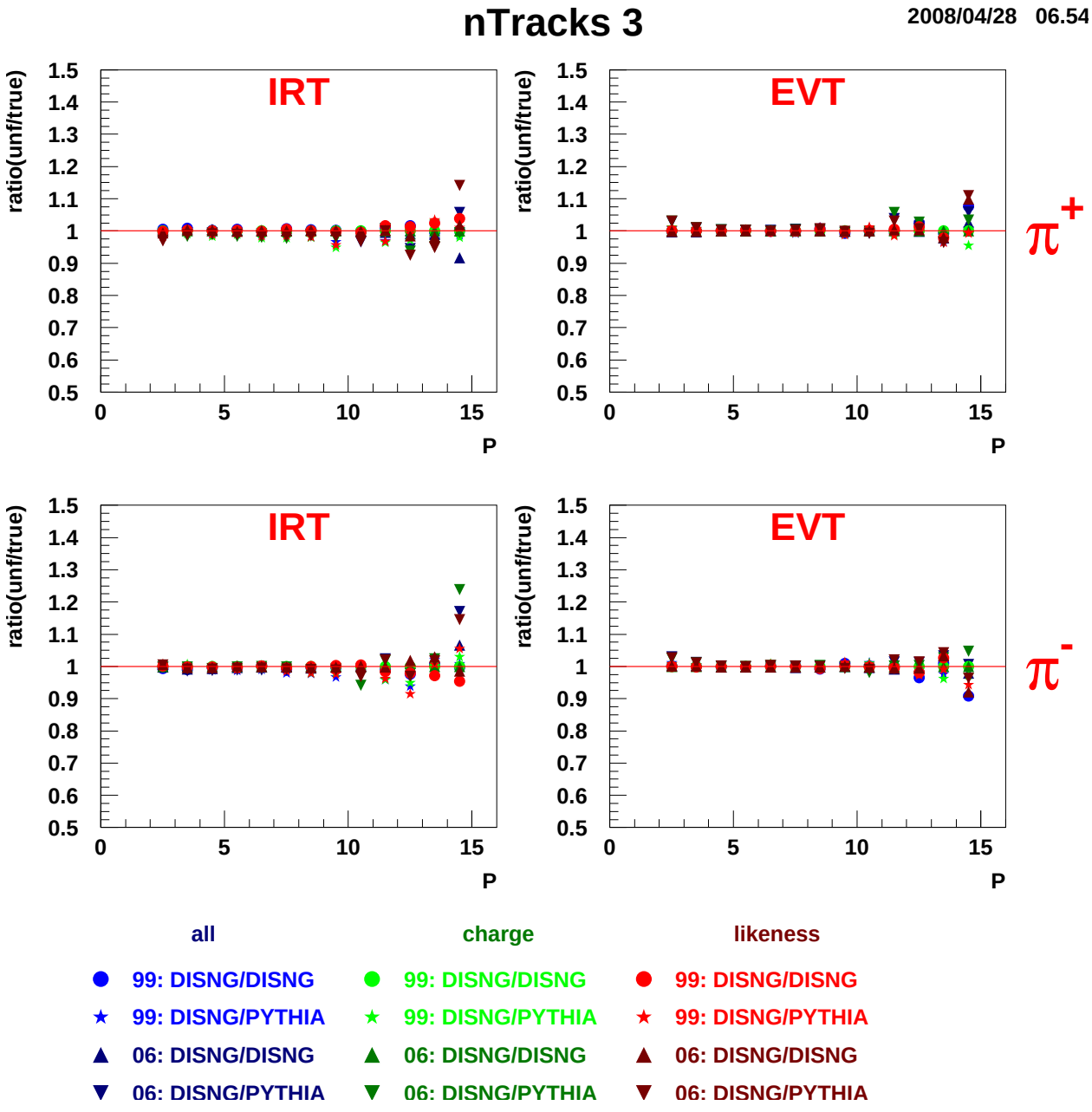

Figure 55: PEPSI Challenge for IRT and EVT for 3+ tracks in the detector half for positive and negative pions. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

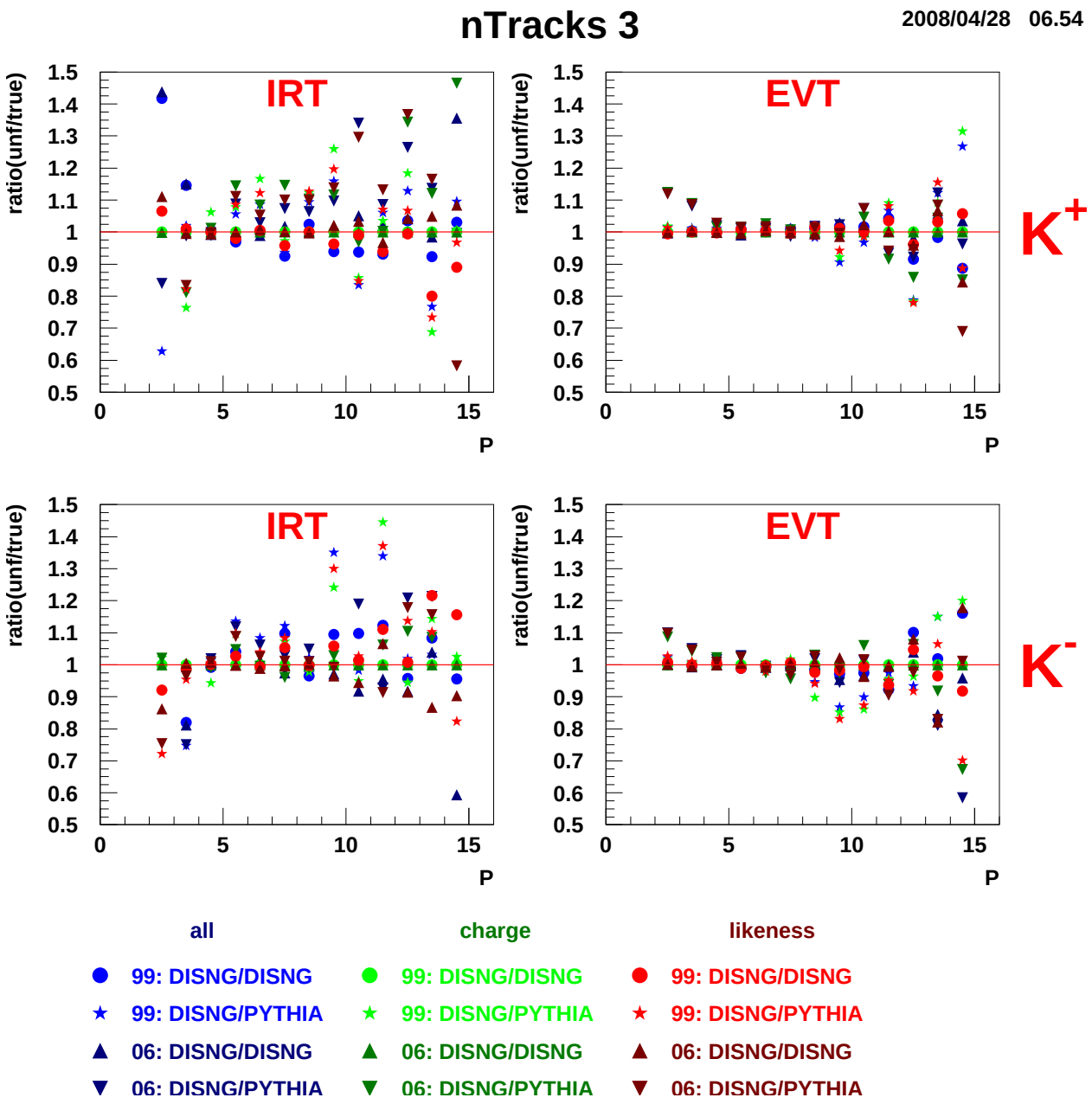

Figure 56: PEPSI Challenge for IRT and EVT for 3+ tracks in the detector half for positive and negative kaons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

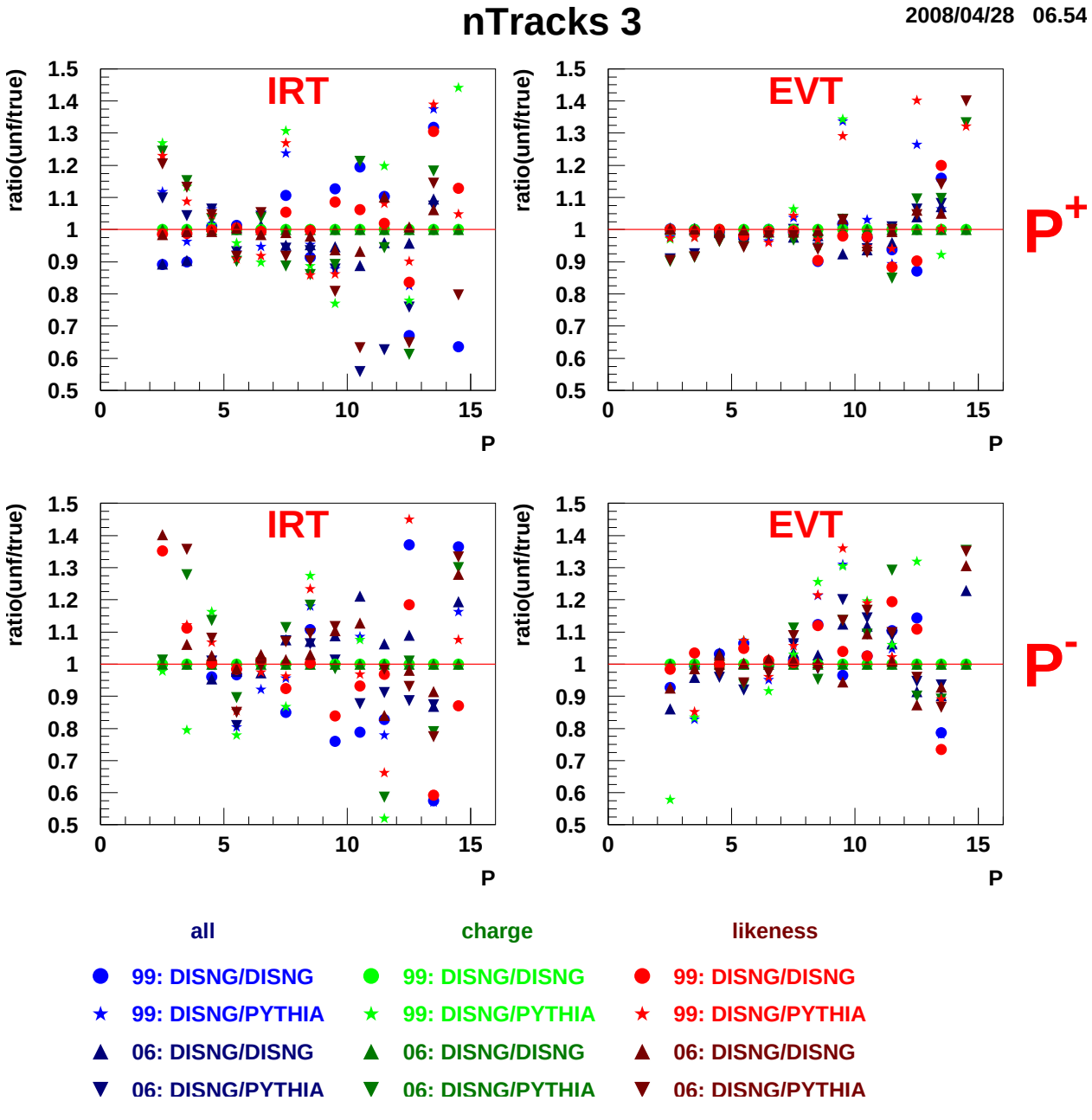

Figure 57: PEPSI Challenge for IRT and EVT for 3+ tracks in the detector half for protons and antiprotons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the production used as (first) the source of the multiplicities and (second) the source of the Pmatrix. All productions use their own background file.

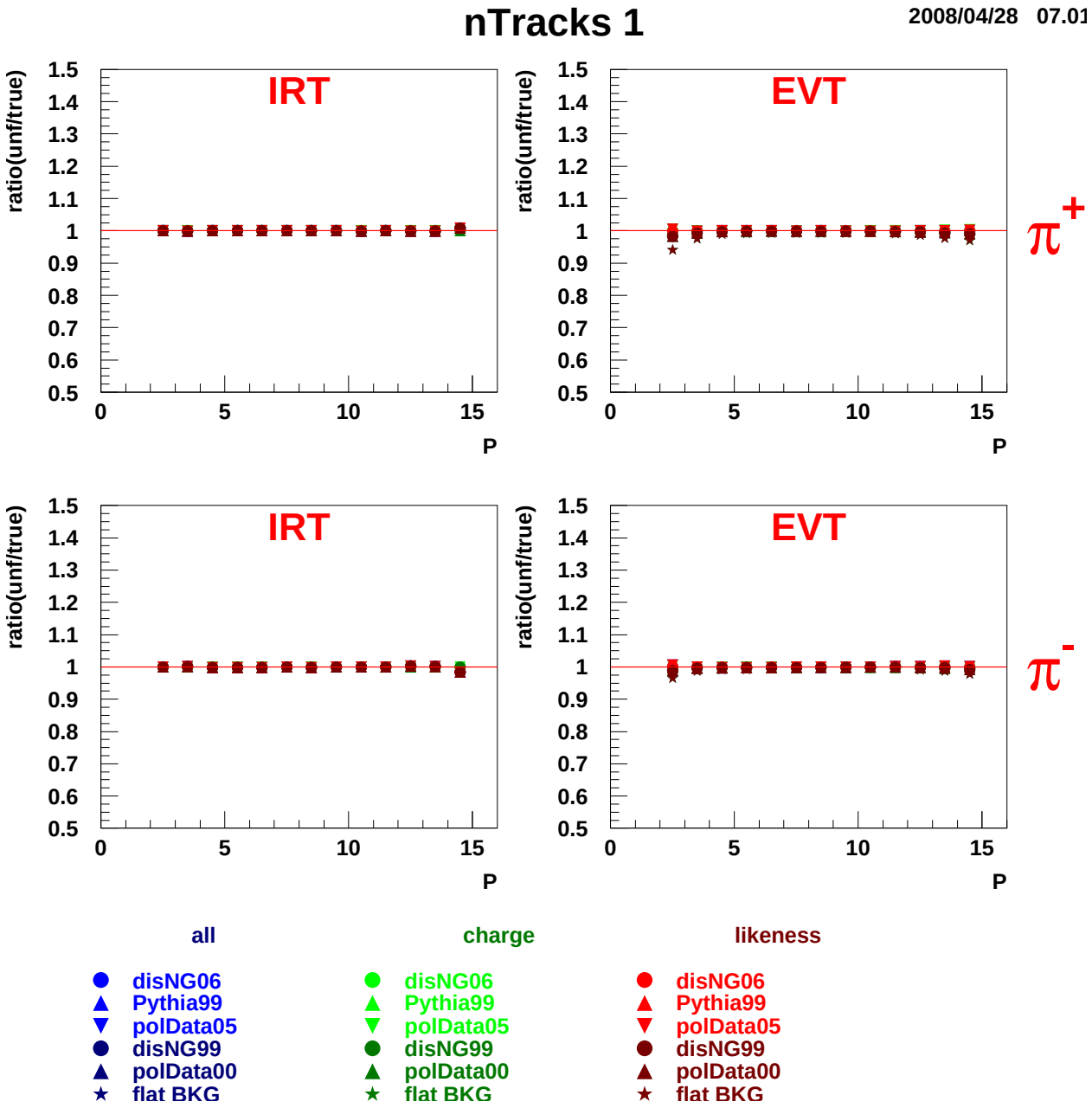

Figure 58: PEPSI Challenge for IRT and EVT (DRT) for 1 track in the detector half for positive and negative pions. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the disNG06 background file, was used to produce the particle multiplicities.

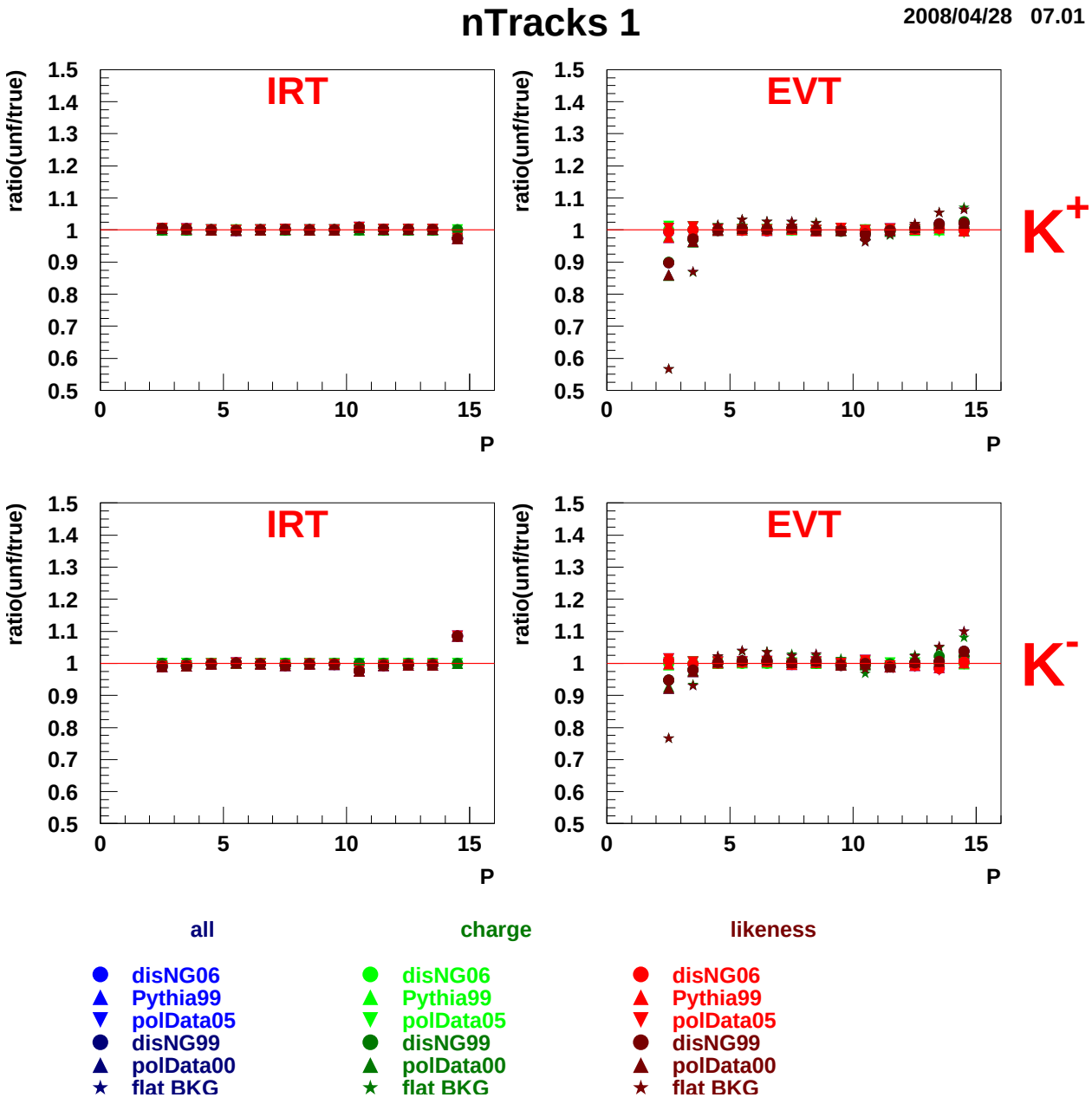

Figure 59: PEPSI Challenge for IRT and EVT (DRT) for 1 track in the detector half for positive and negative kaons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the DISNG06 background file, was used to produce the particle multiplicities.

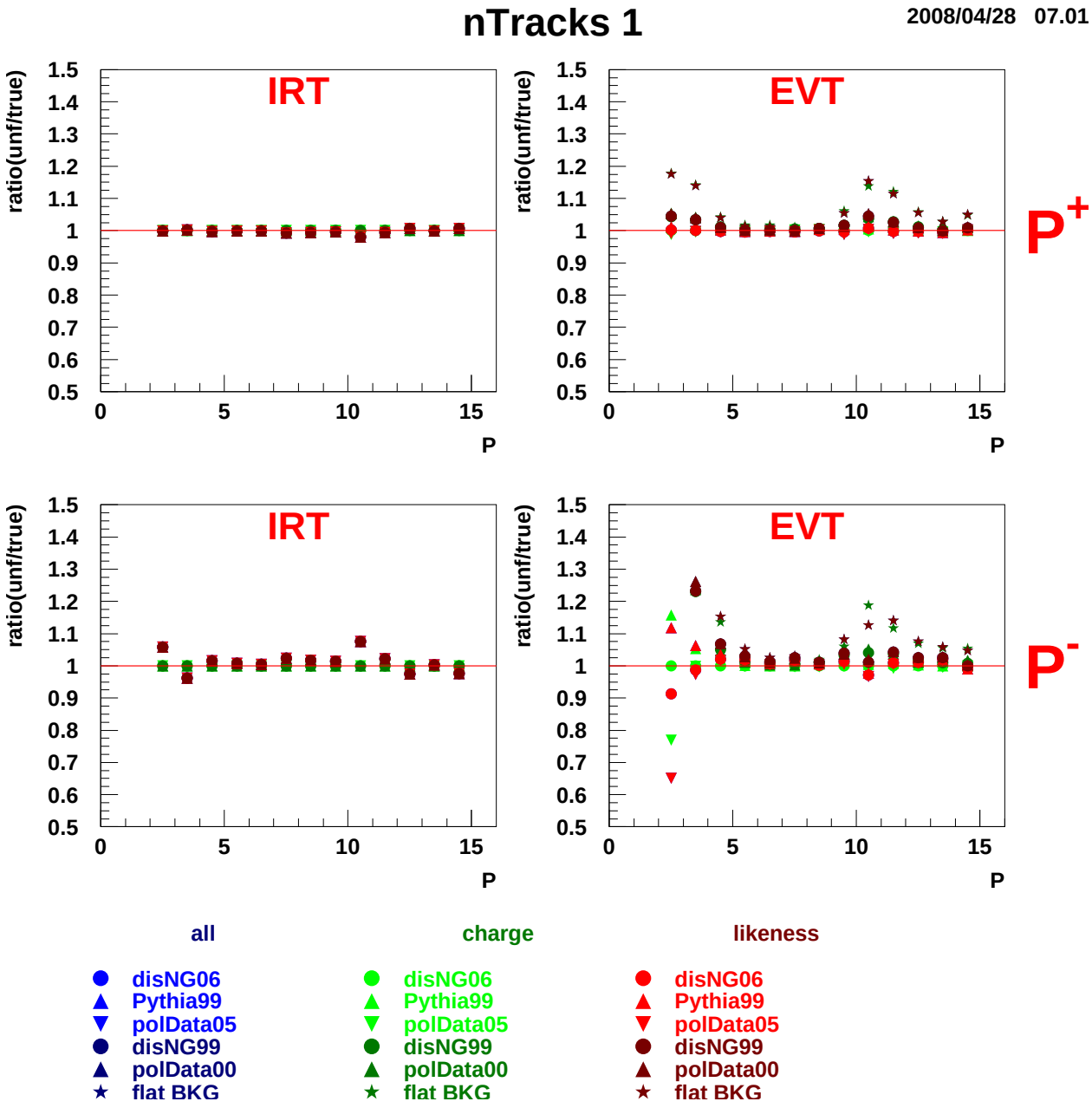

Figure 60: PEPSI Challenge for IRT and EVT (DRT) for 1 track in the detector half for protons and anti-protons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the DISNG06 background file, was used to produce the particle multiplicities.

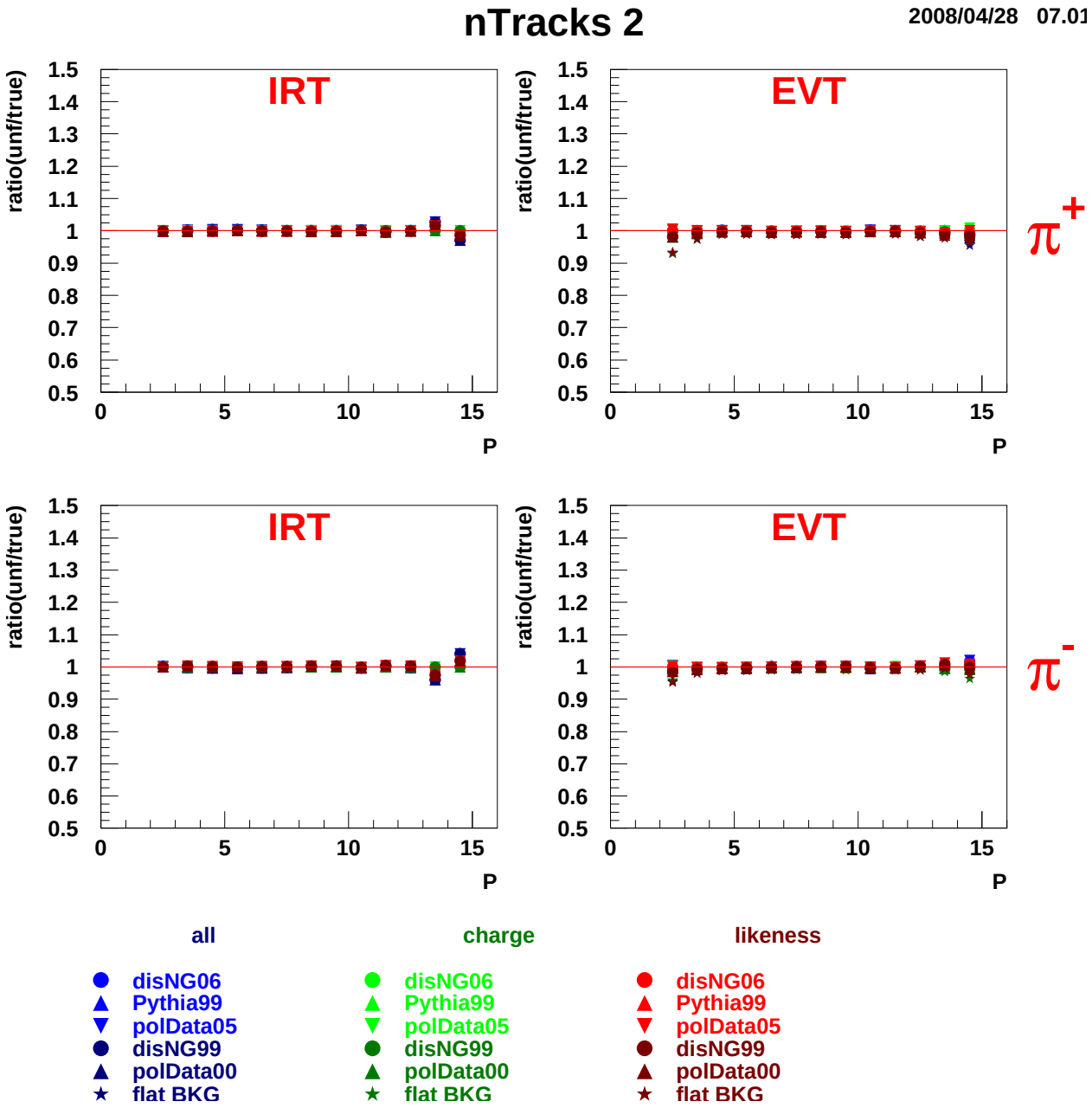

Figure 61: PEPSI Challenge for IRT and EVT for 2 tracks in the detector half for positive and negative pions. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the DISNG06 background file, was used to produce the particle multiplicities.

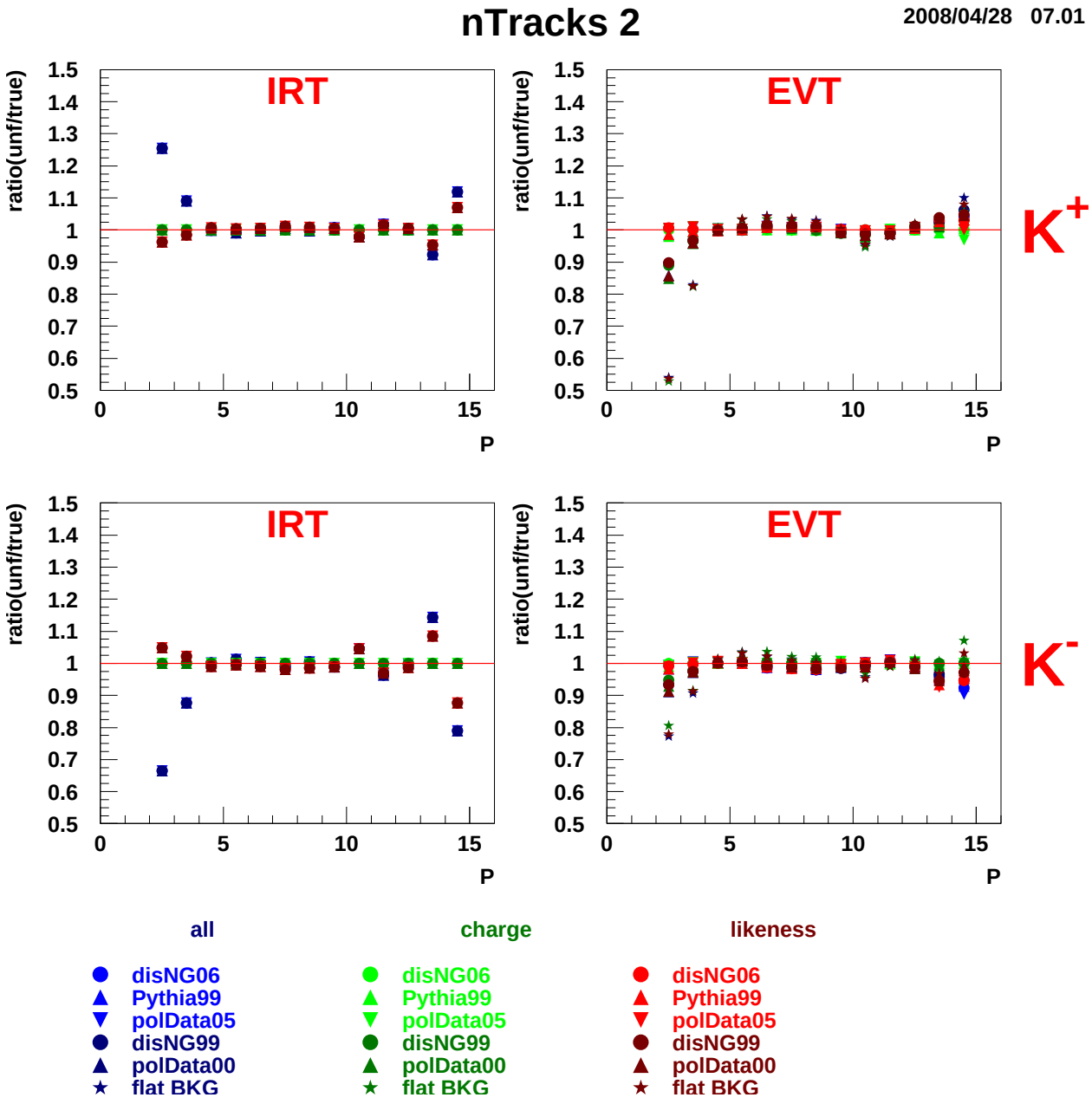

Figure 62: PEPSI Challenge for IRT and EVT for 2 tracks in the detector half for positive and negative kaons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the disNG06 background file, was used to produce the particle multiplicities.

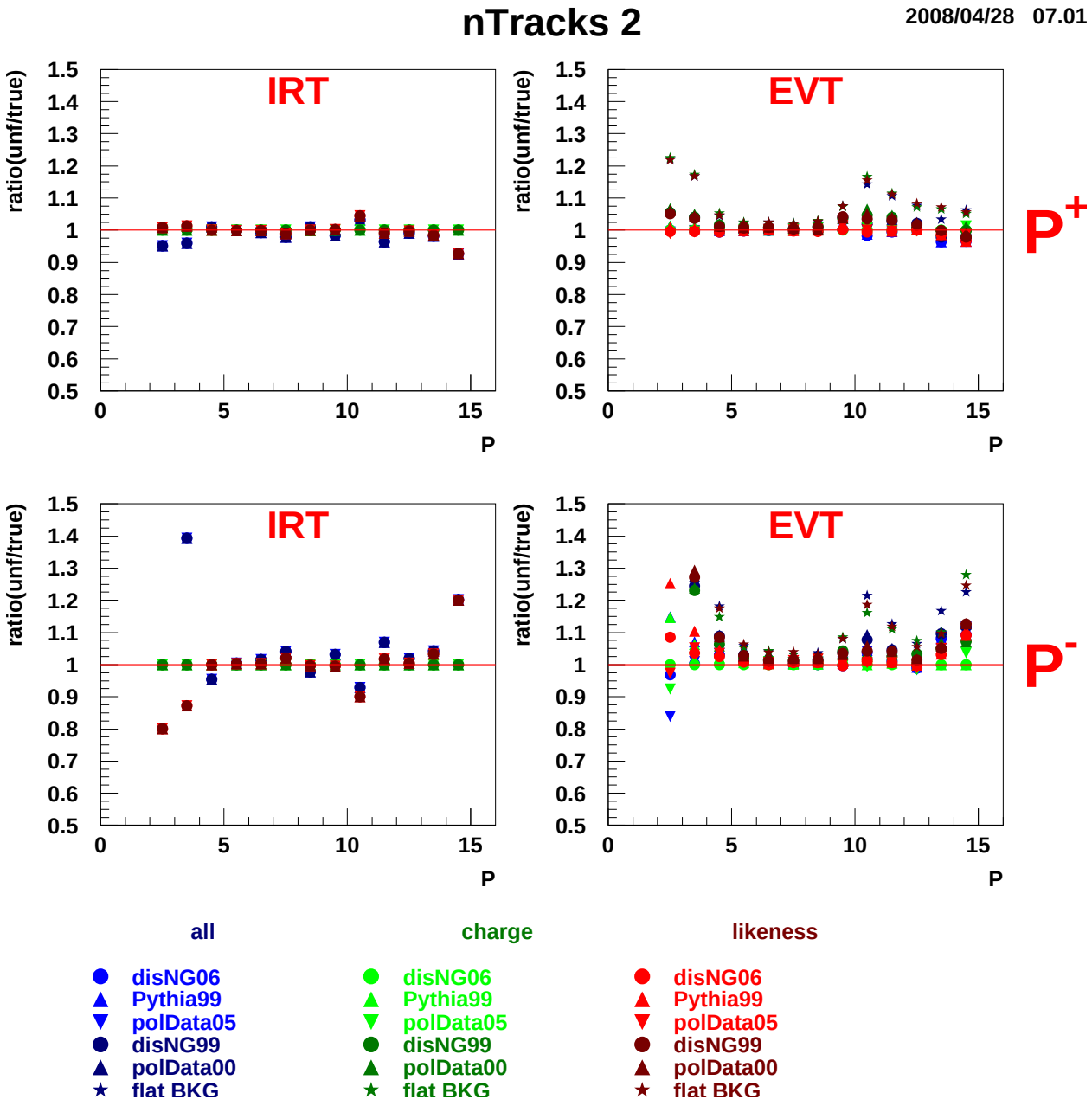

Figure 63: PEPSI Challenge for IRT and EVT for 2 tracks in the detector half for protons and anti-protons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the disNG06 background file, was used to produce the particle multiplicities.

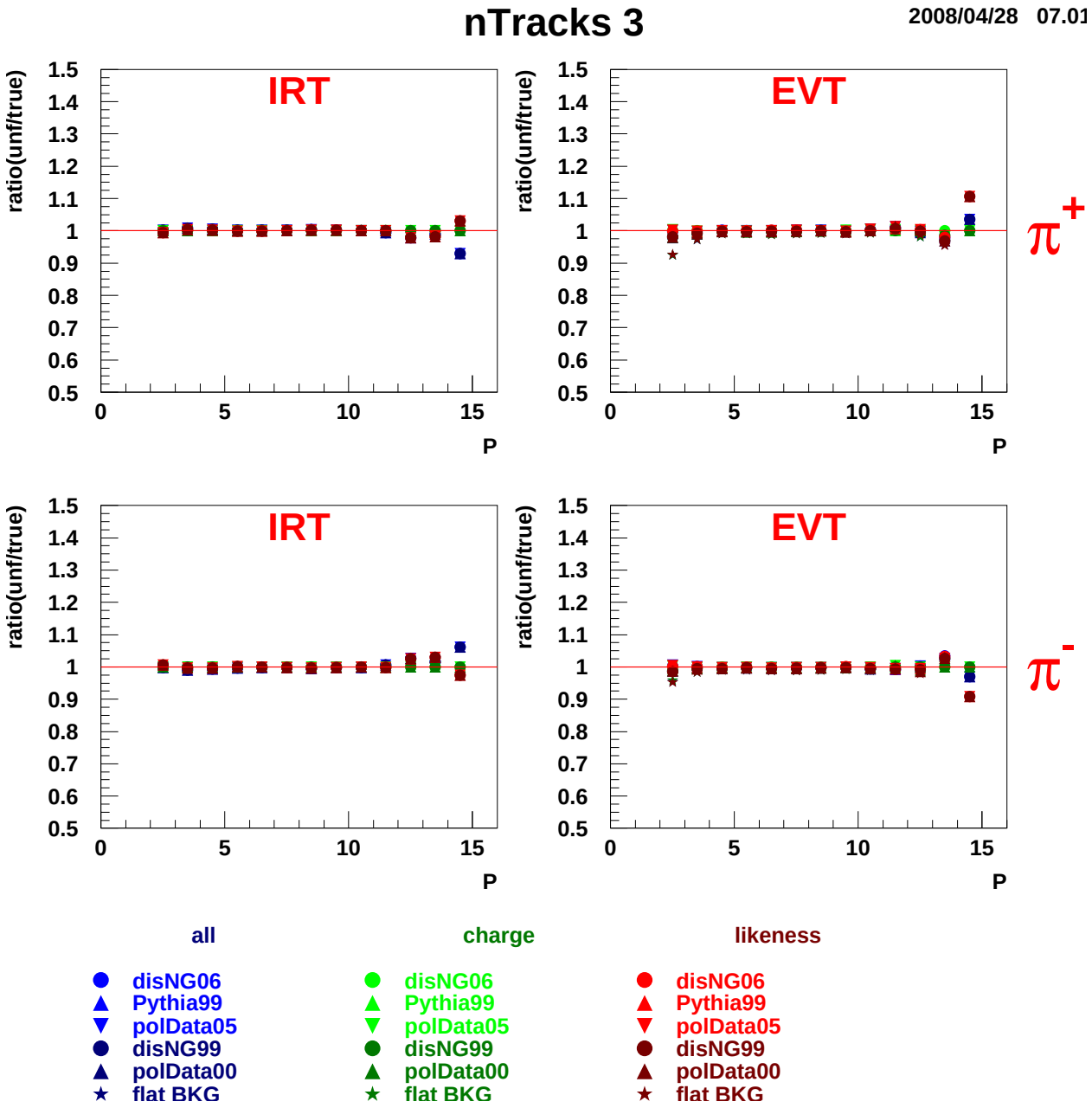

Figure 64: PEPSI Challenge for IRT and EVT for 3+ tracks in the detector half for positive and negative pions. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the disNG06 background file, was used to produce the particle multiplicities.

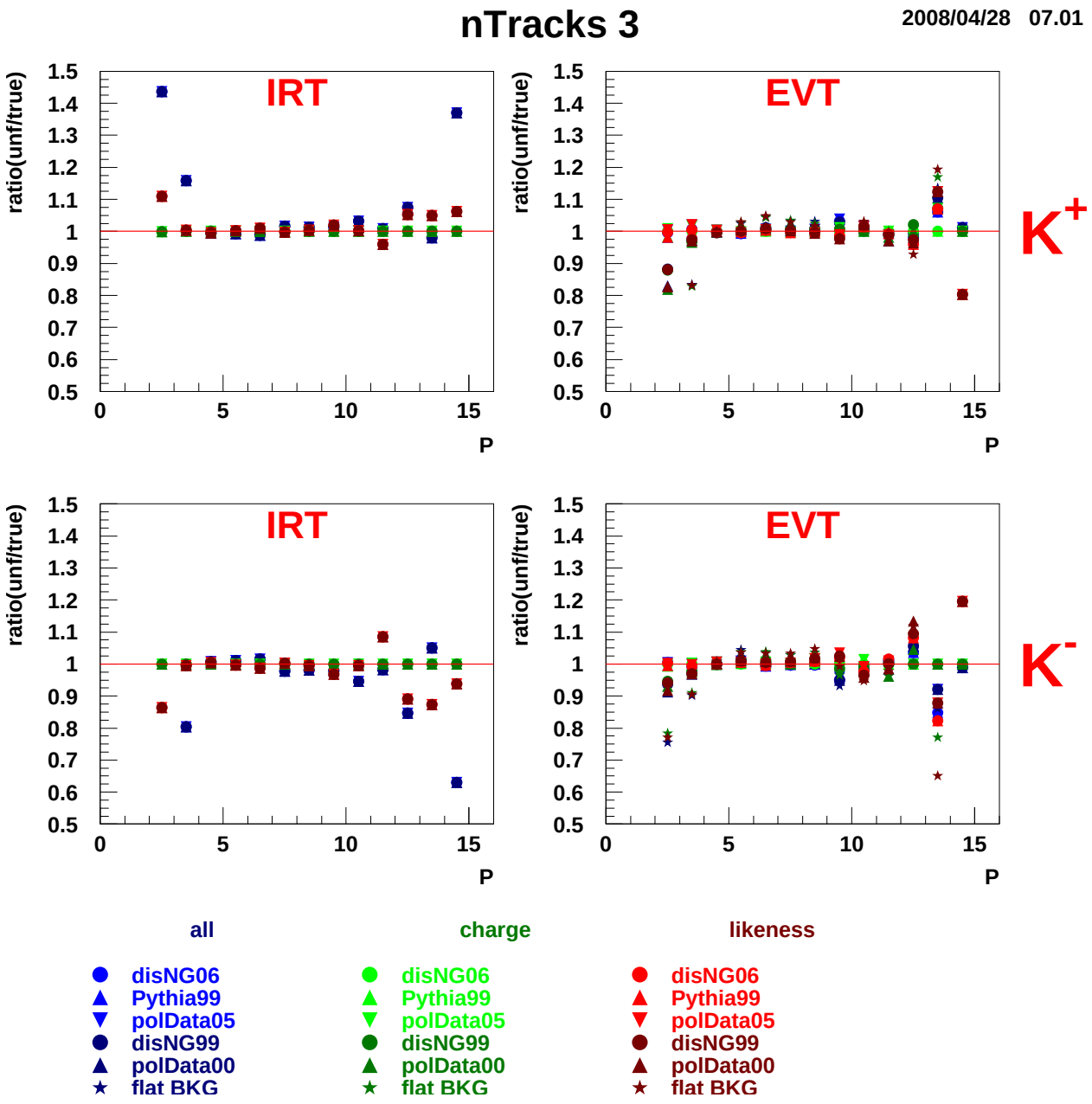

Figure 65: PEPSI Challenge for IRT and EVT for 3+ tracks in the detector half for positive and negative kaons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the disNG06 background file, was used to produce the particle multiplicities.

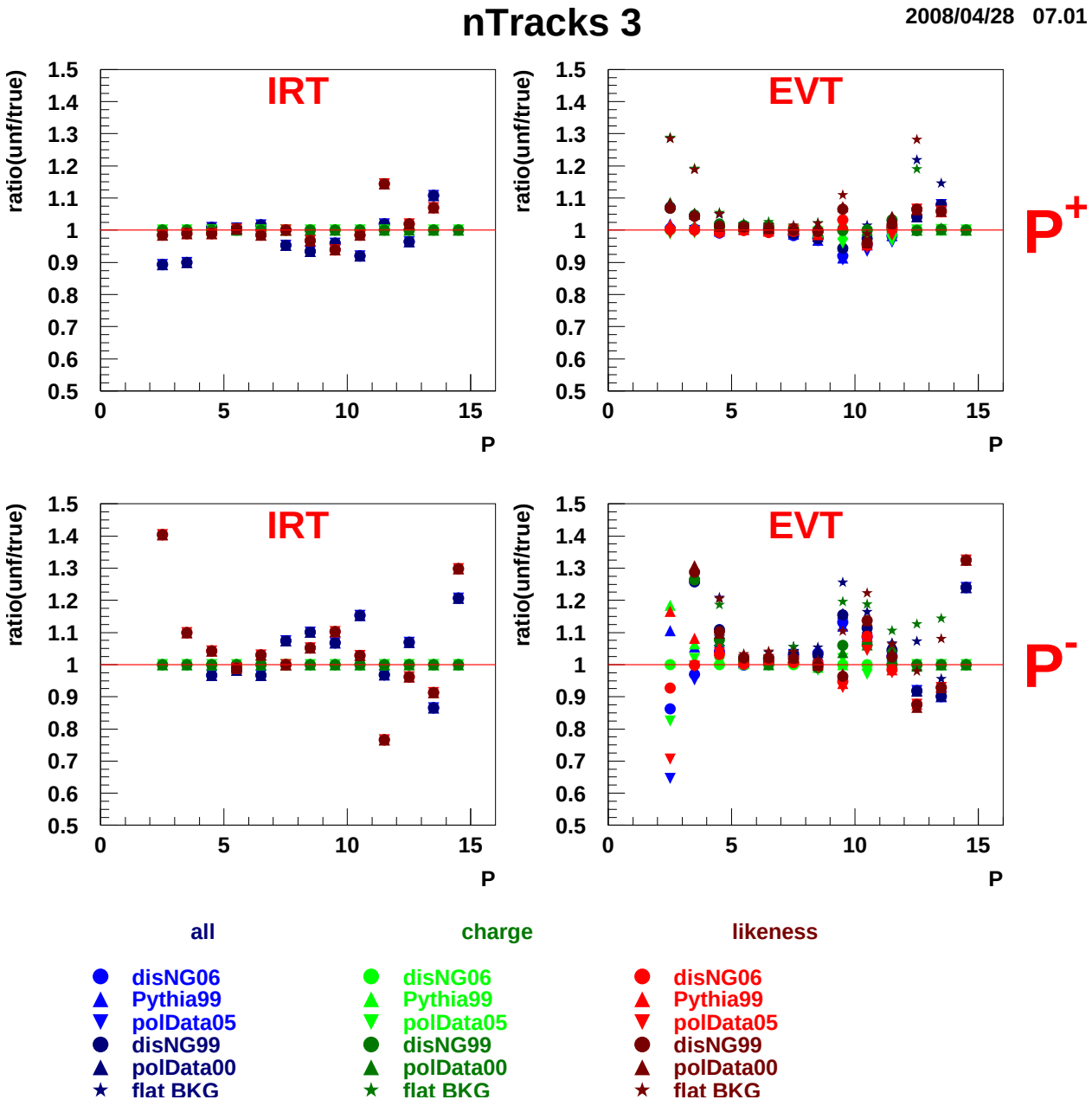

Figure 66: PEPSI Challenge for IRT and EVT for 3+ tracks in the detector half for protons and anti-protons. The colors indicate the type of Pmatrix used (all charges together, charge separated, or likeness separated). The symbols denote the background assumption in the production used to extract the Pmatrices. In all cases the same disNG production with 2006 geometry, which uses the disNG06 background file, was used to produce the particle multiplicities.

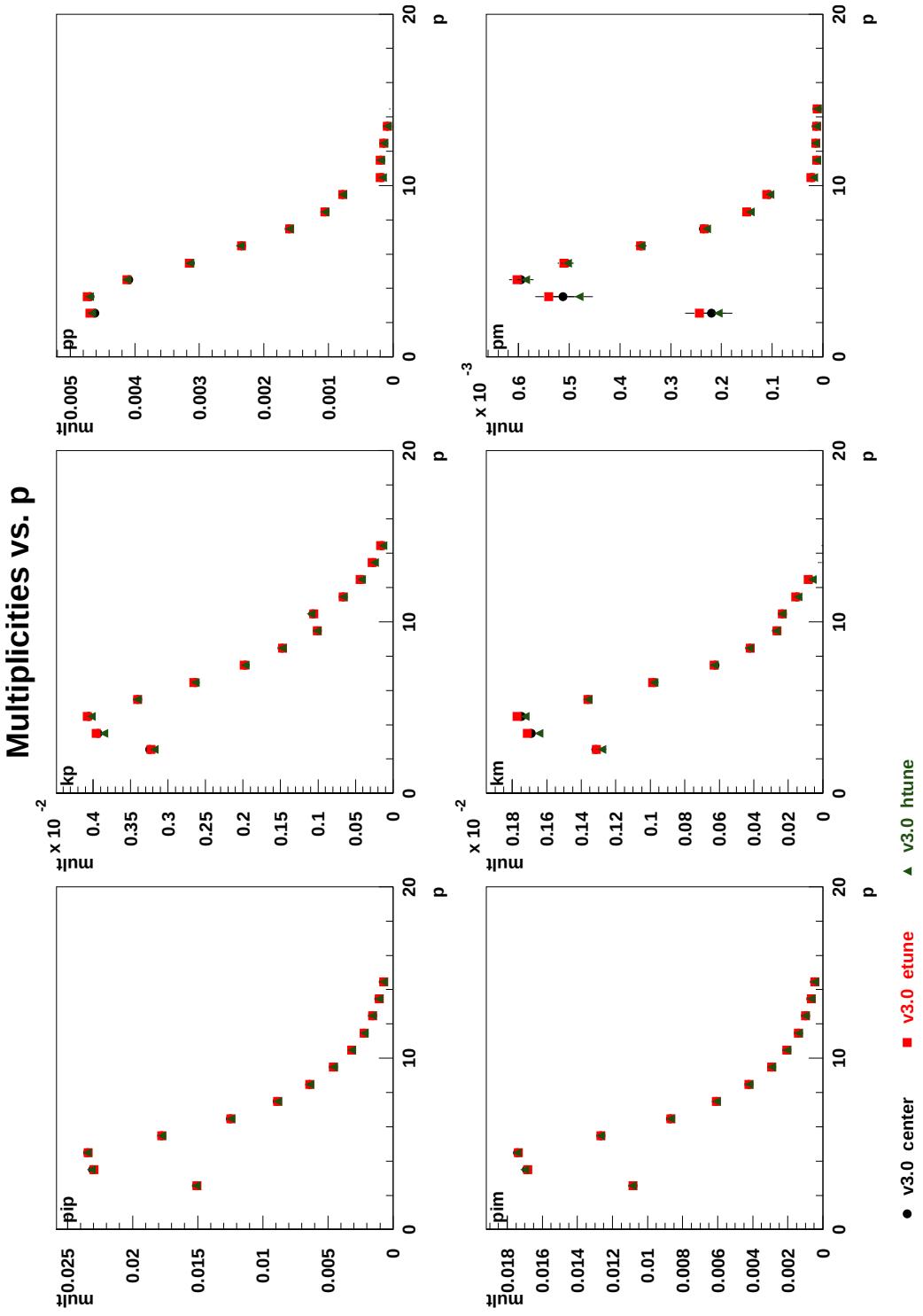

Figure 67: Comparison of the 00d2 multiplicities unfolded with the version 3.0 (99 geometry, charge combined), e-, h-, and c-tune IRT Pmatrices, as a function of momentum. In the top row are shown the positive pion, kaon and proton multiplicities. In the bottom row are the respective negative particles.

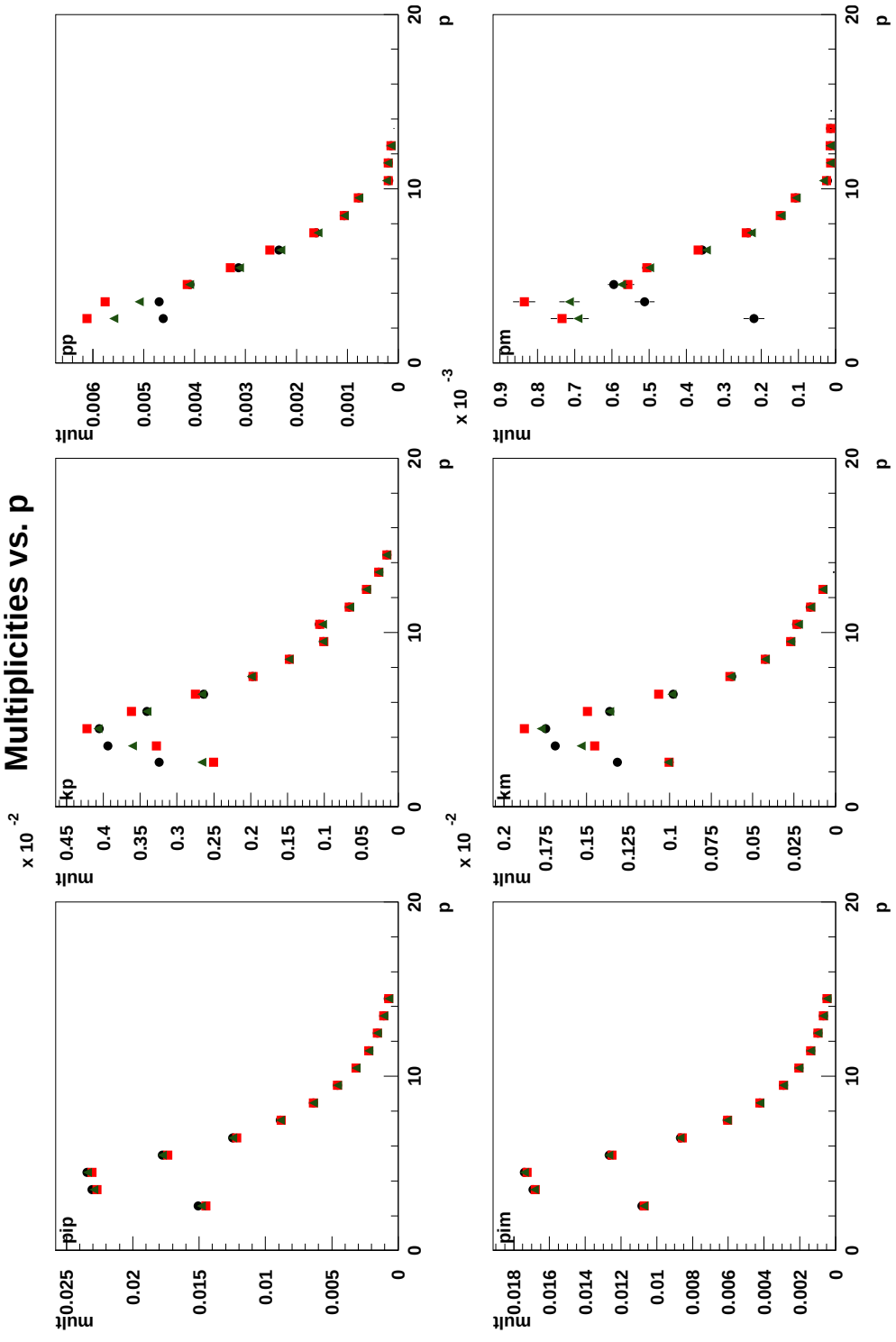

Figure 68: Comparison of the 00d2 multiplicities unfolded with various IRT Pmatrices (see Figure 33), as a function of momentum. In the top row are shown the positive pion, kaon and proton multiplicities. In the bottom row are the respective negative particles.

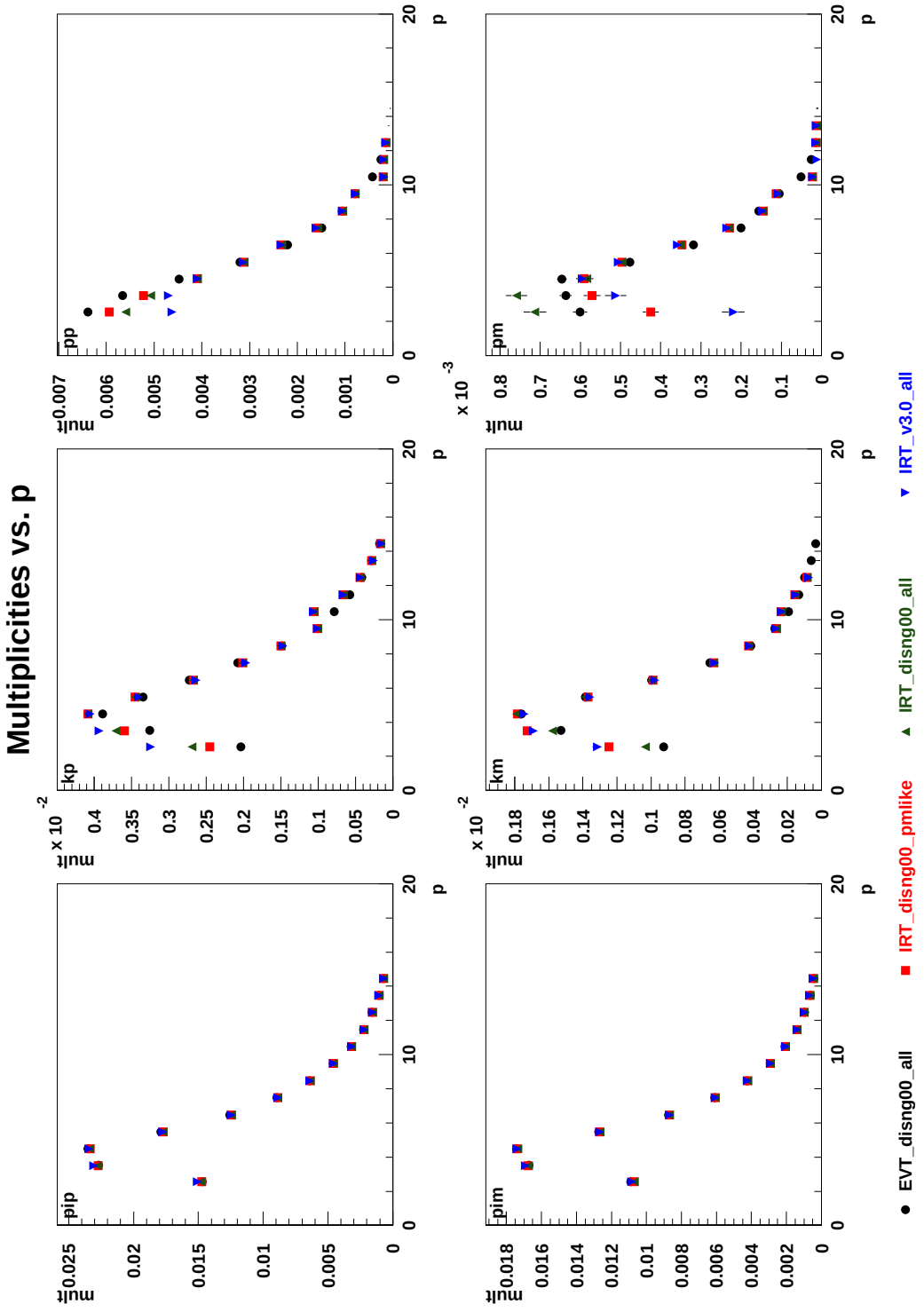

Figure 69: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using various IRT and EVT Pmatrices, all produced using their own respective backgrounds.

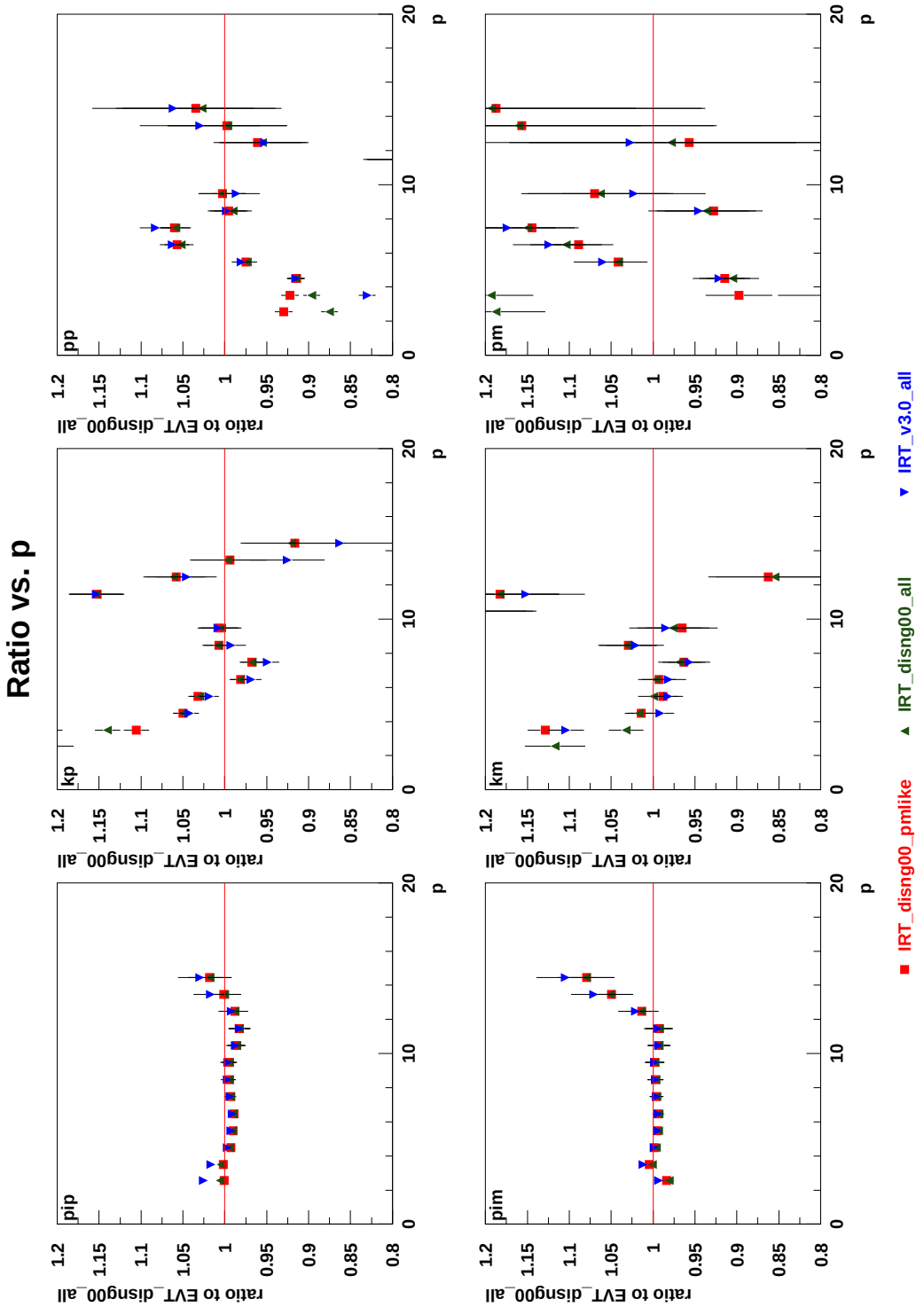

Figure 70: Ratio of the multiplicities shown in Figure 69.

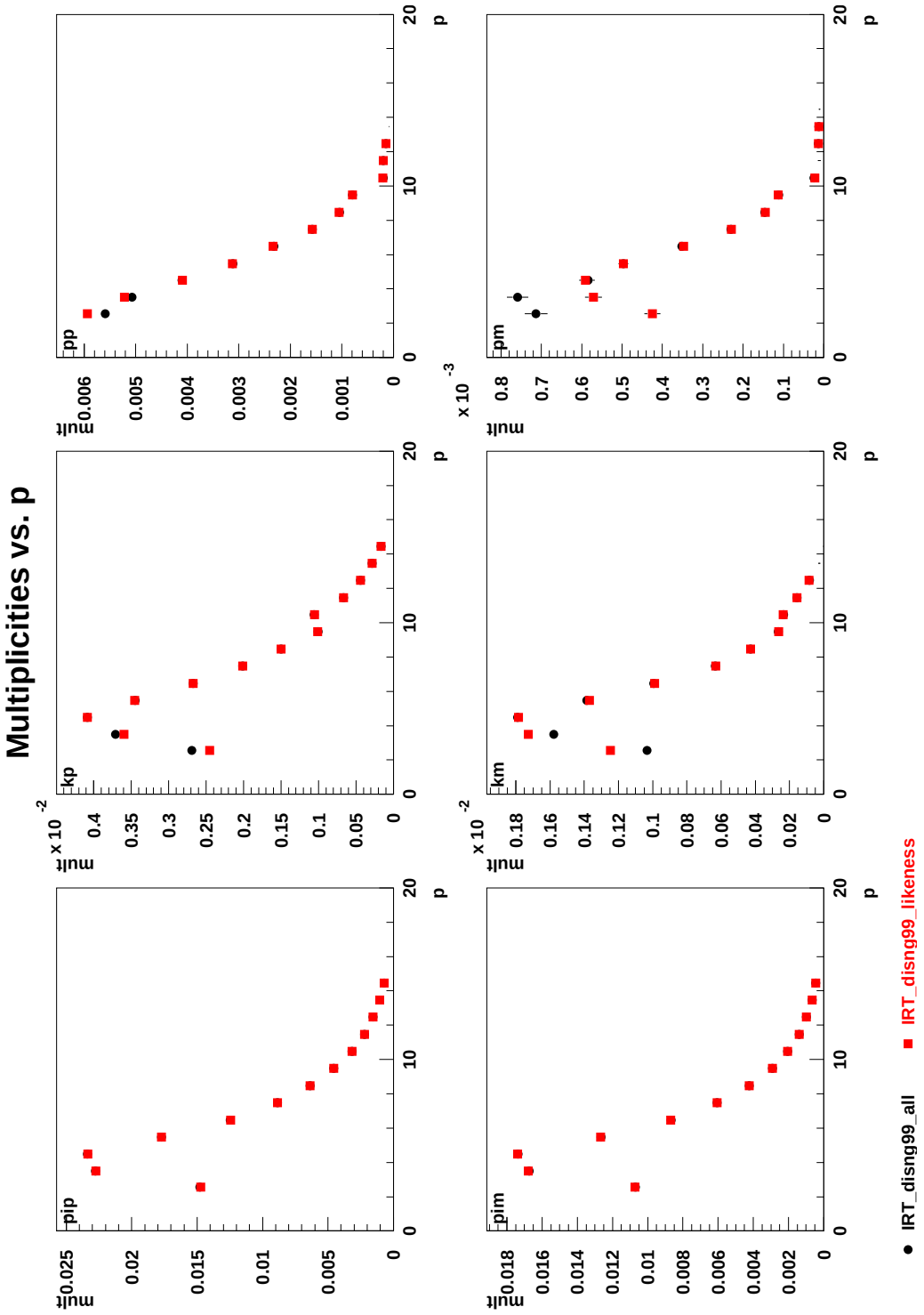

Figure 71: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using IRT Pmatrices from disNG 1999 geometry that are charge combined (all) or likeness separated.

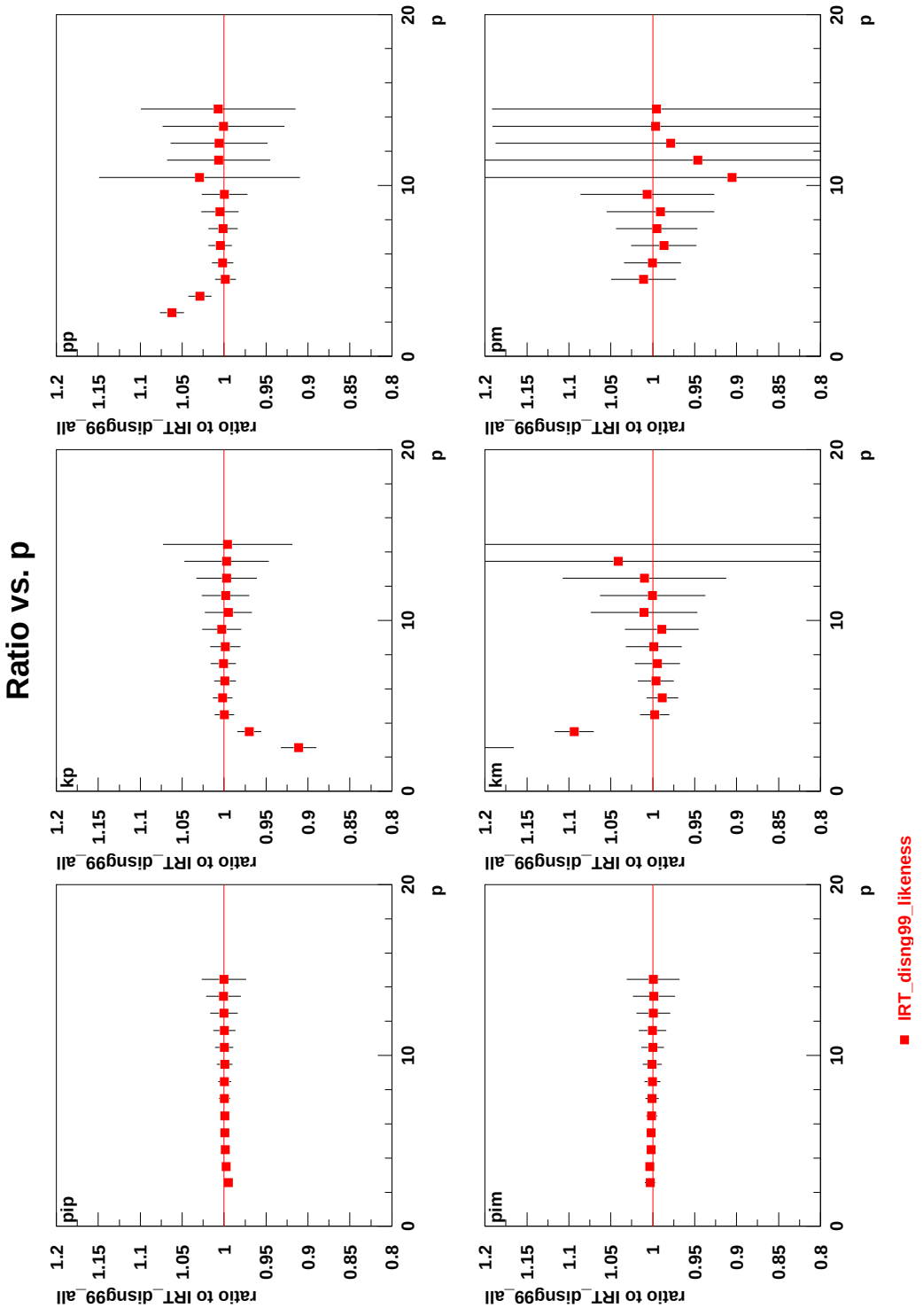

Figure 72: Ratio of the multiplicities shown in Figure 71.

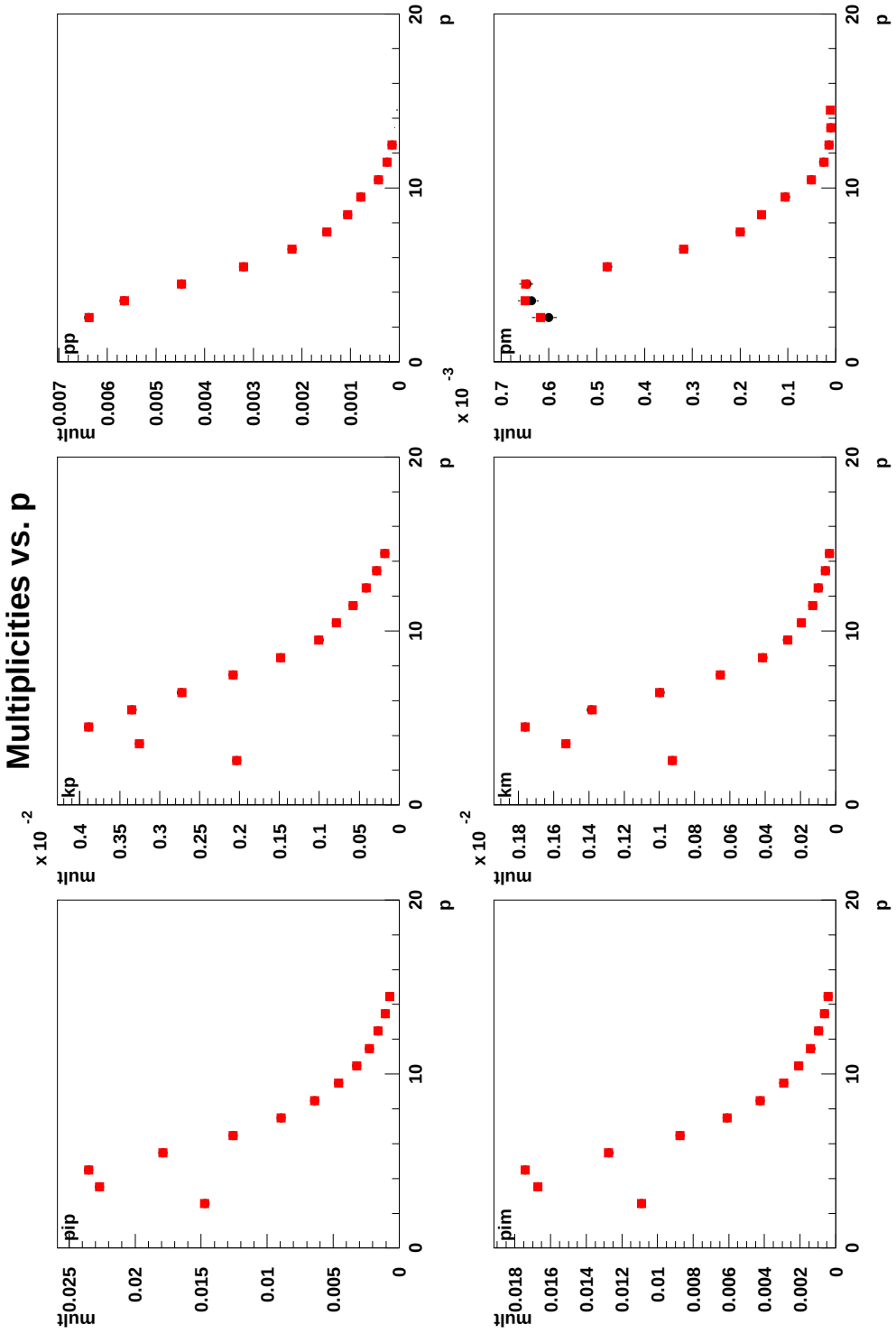

Figure 73: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using EVT Pmatrices from disNG 1999 geometry with the disNG99 background file that are charge combined (all) and likeness separated.

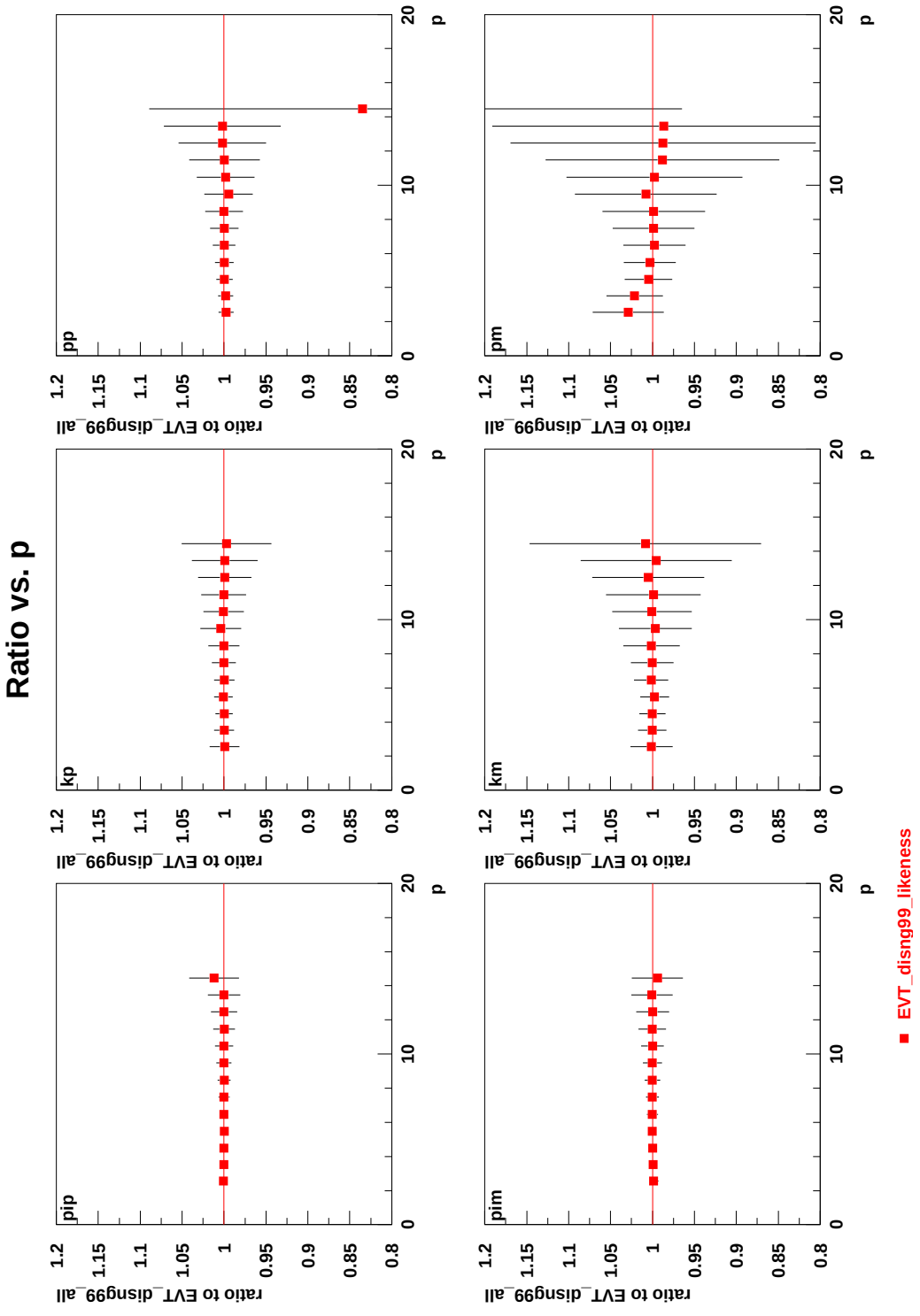

Figure 74: Ratio of the multiplicities shown in Figure 73.

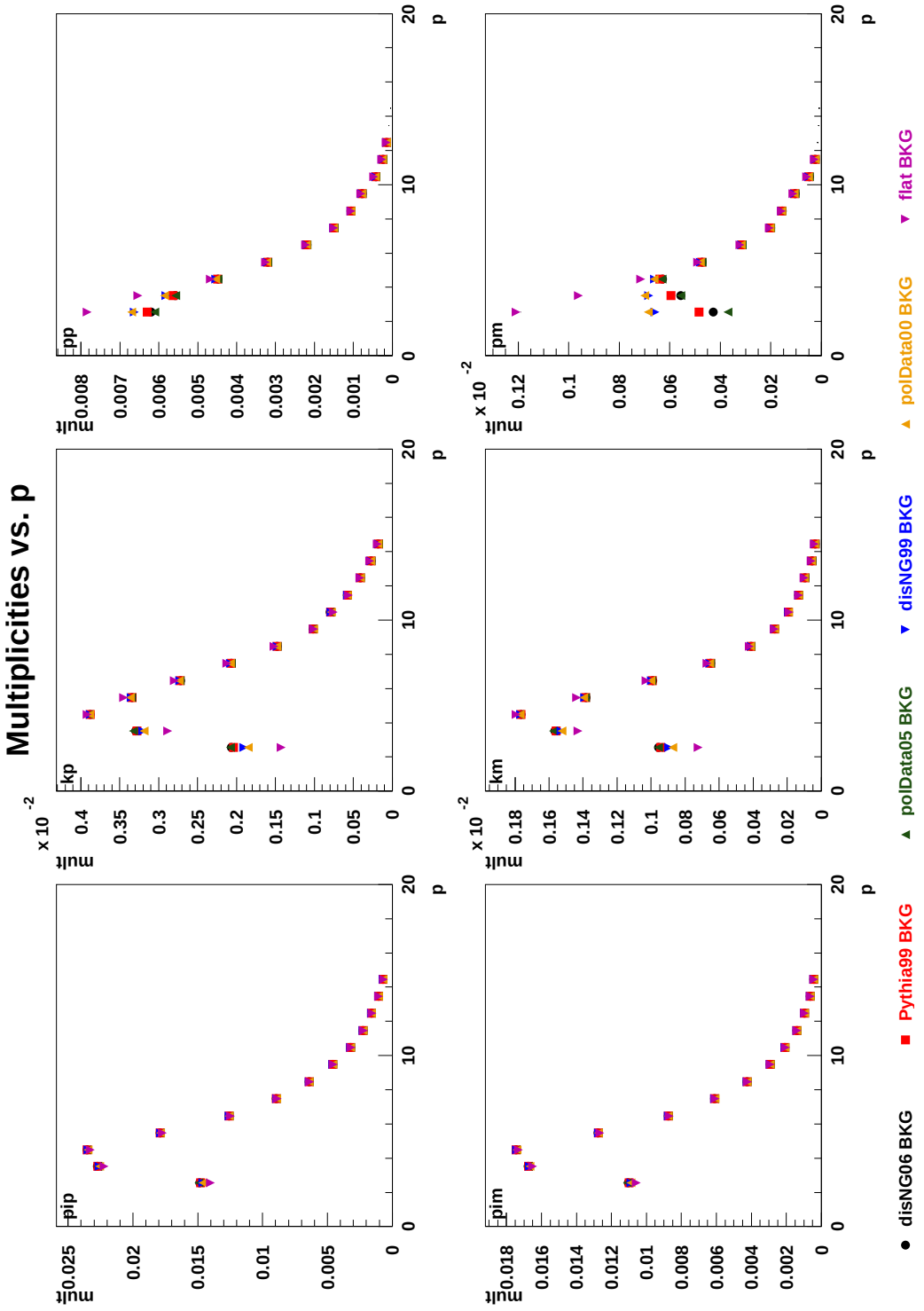

Figure 75: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using charge combined EVT Pmatrices extracted from a disNG with 2006 geometry production using different background assumptions (see section 6.4).

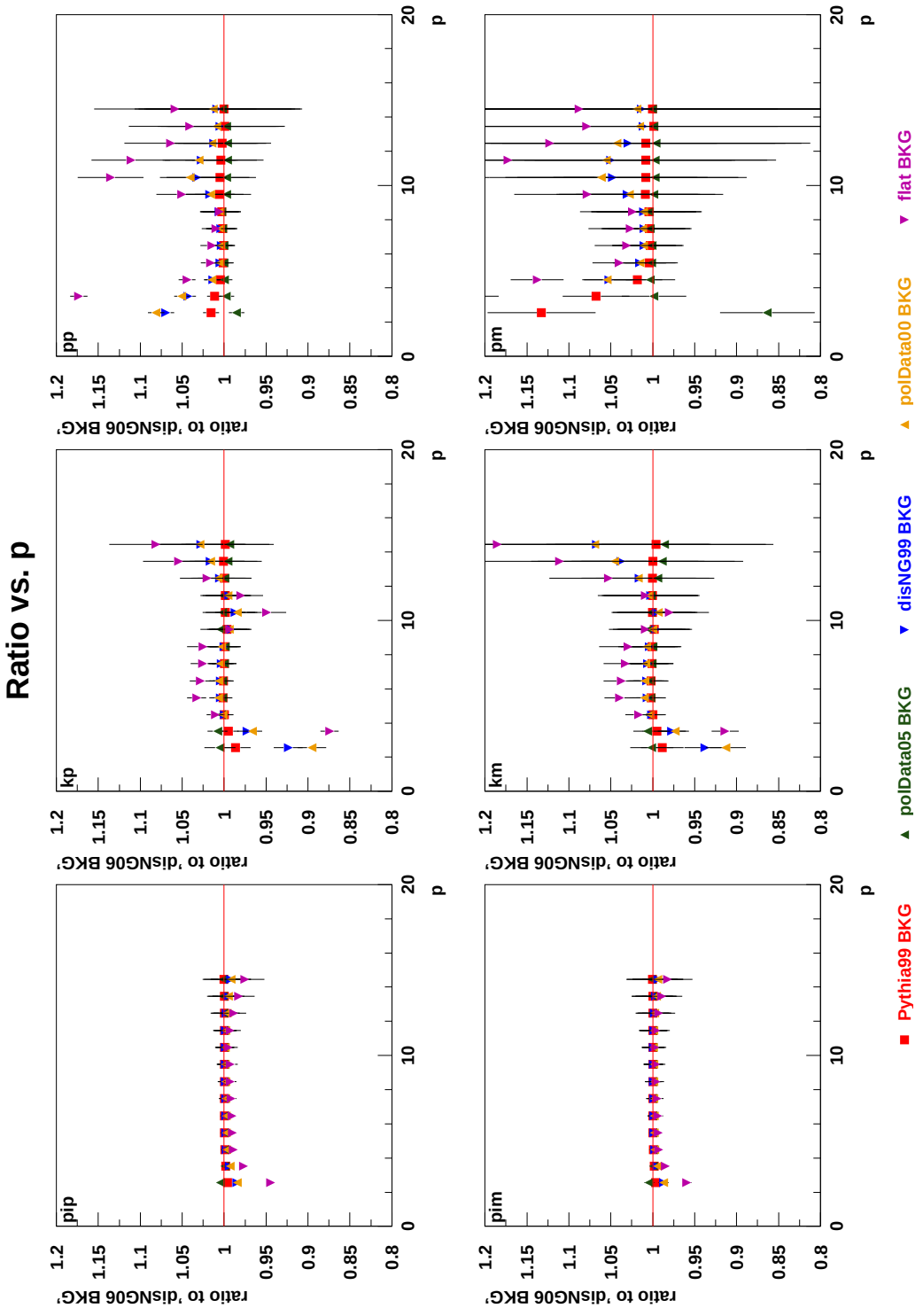

Figure 76: Ratios of the multiplicities shown in Figure 75, relative to the first data set.

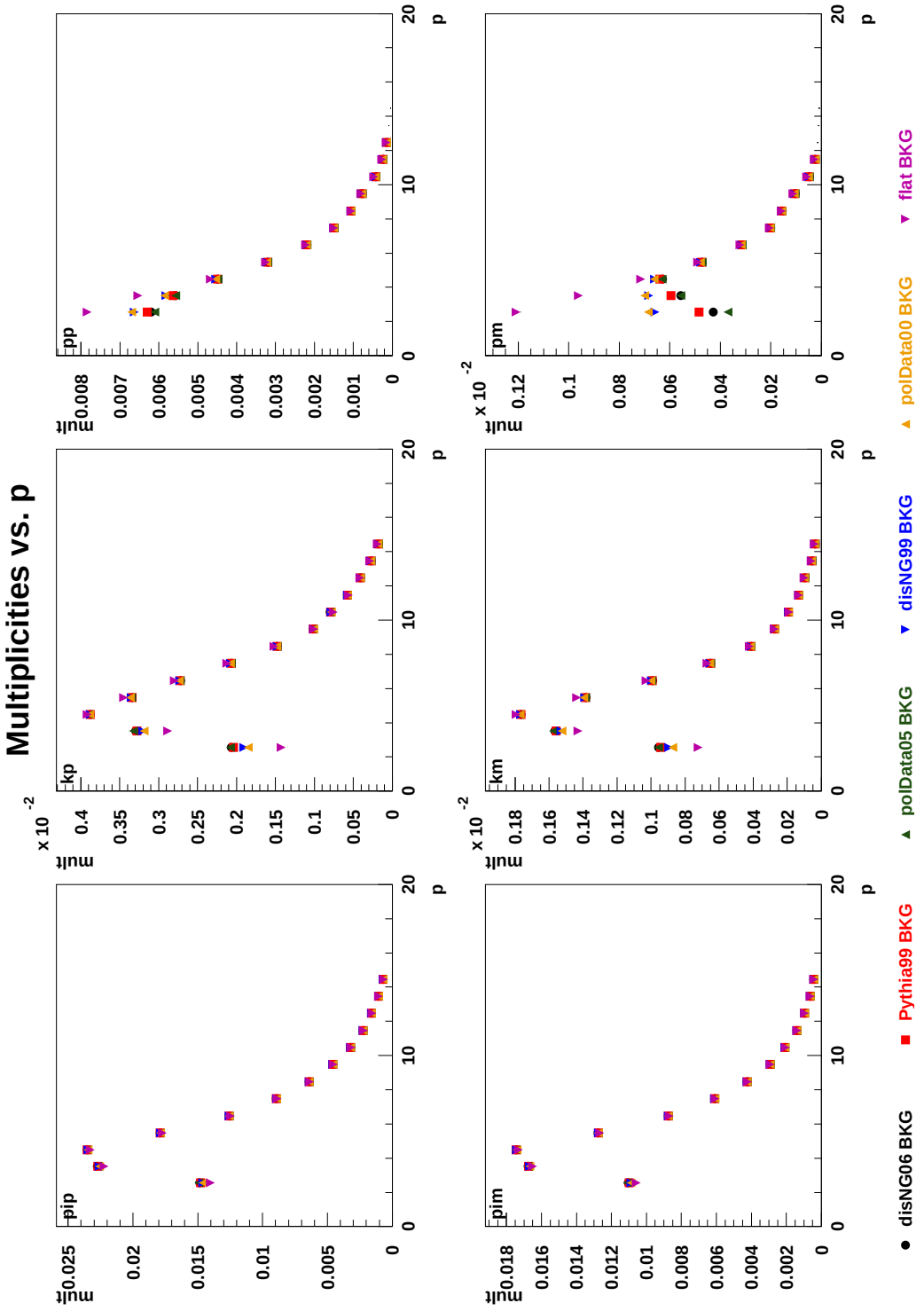

Figure 77: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using likeness separated EVT Pmatrices extracted from disNG 2006 using different background assumptions (see section 6.4).

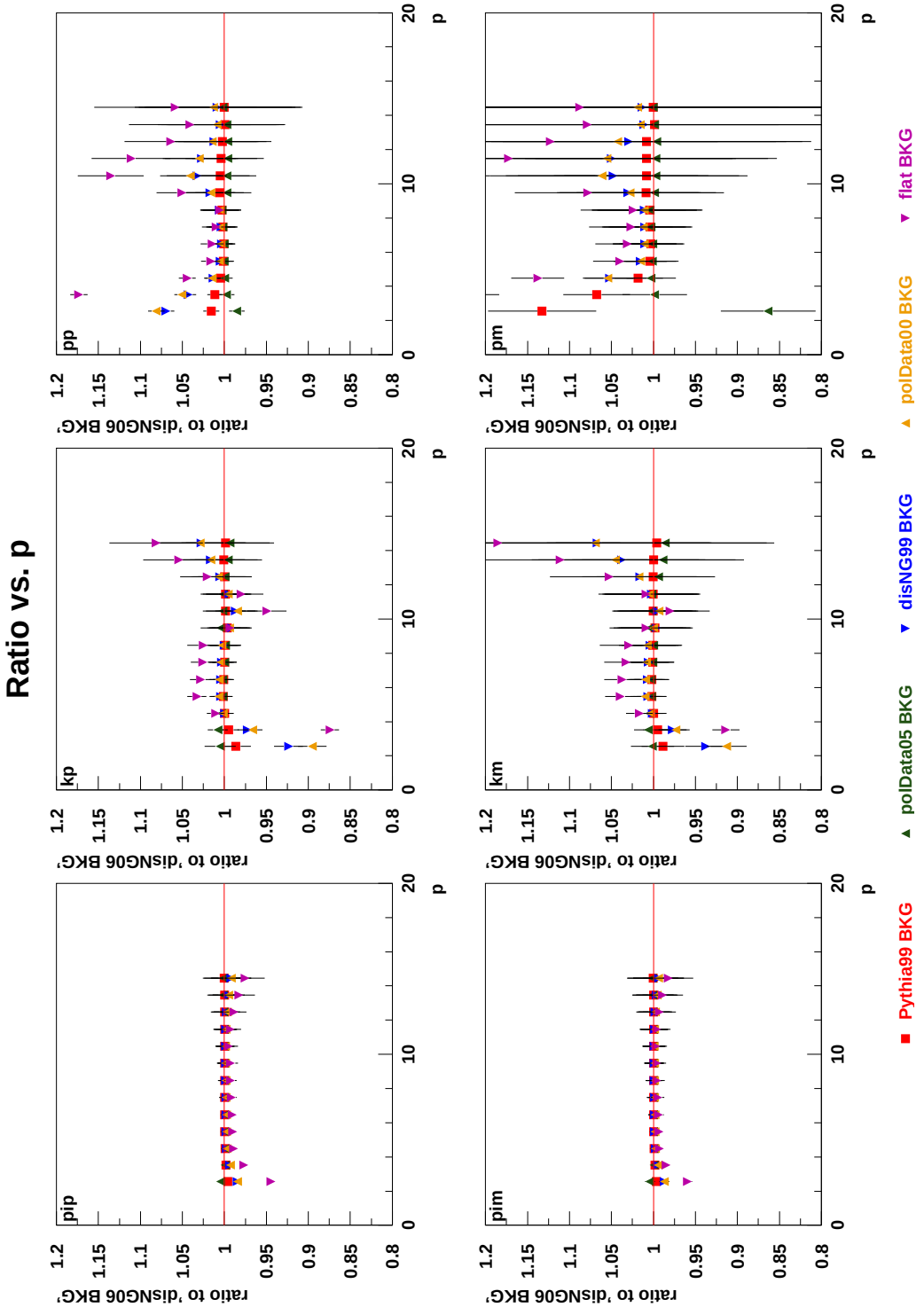

Figure 78: Ratios of the multiplicities shown in Figure 77, relative to the first data set.

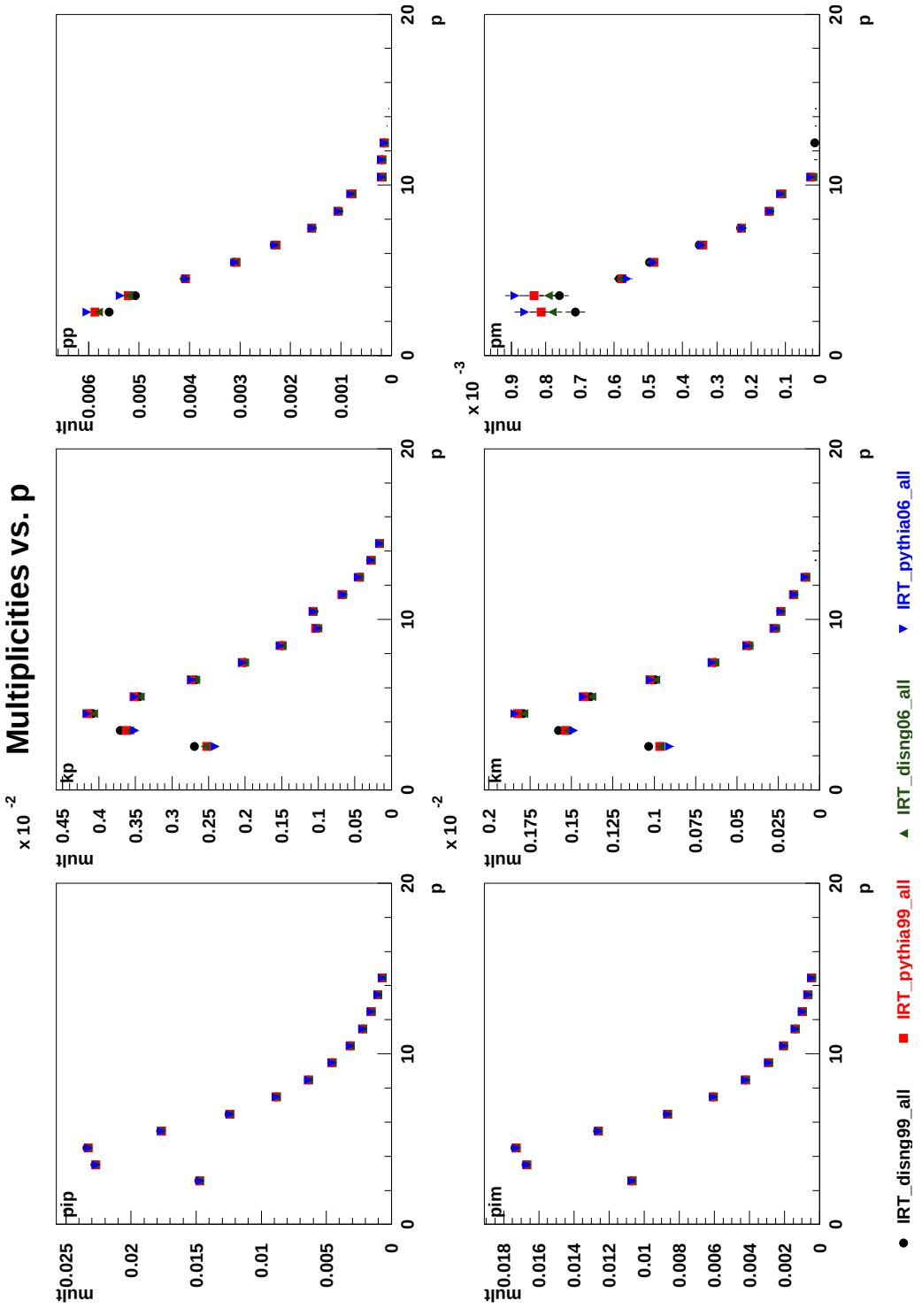

Figure 79: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using charge combined IRT Pmatrices extracted from MC productions using different generators and geometries.

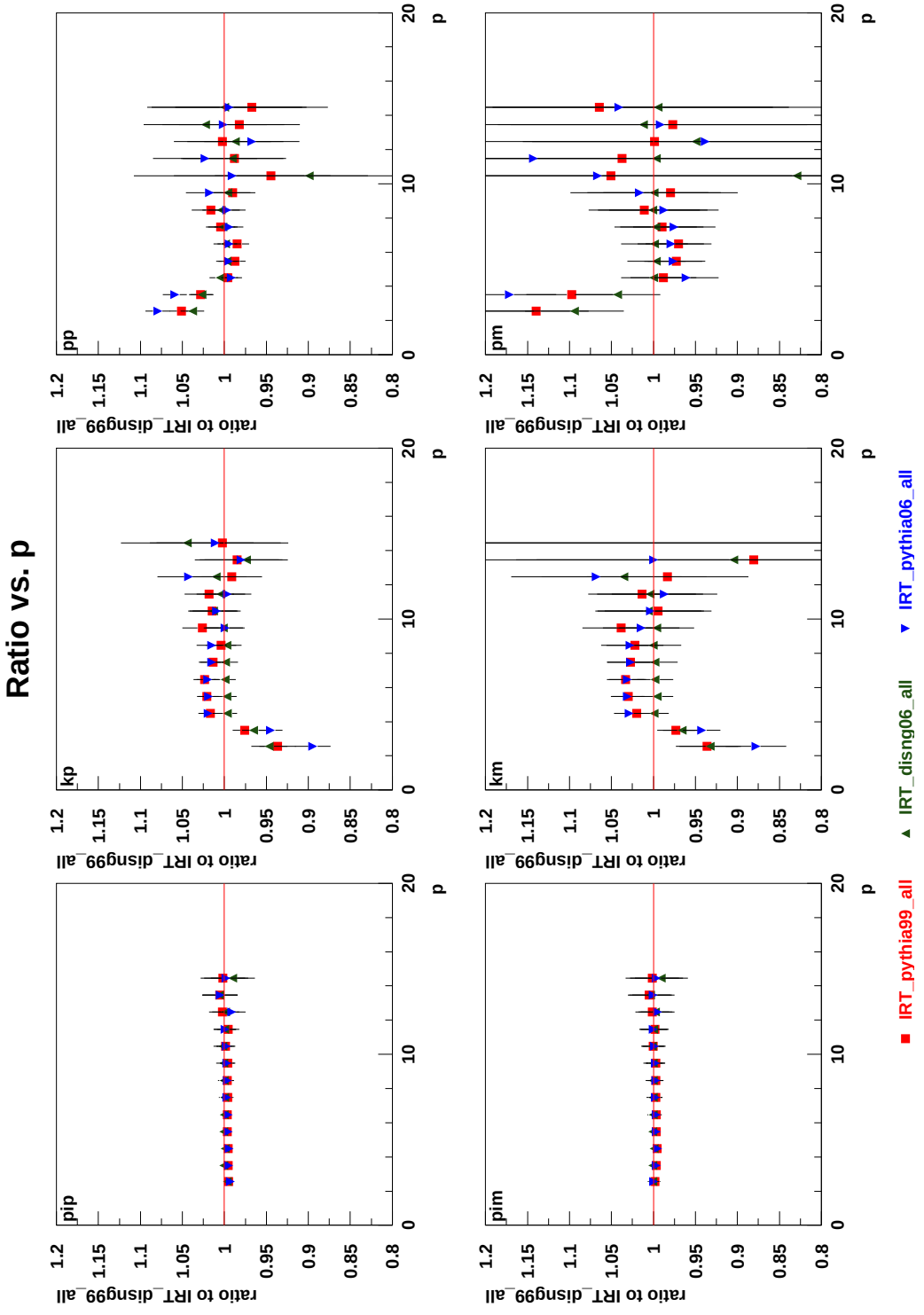

Figure 80: Ratios of the multiplicities shown in Figure 79, relative to the first data set.

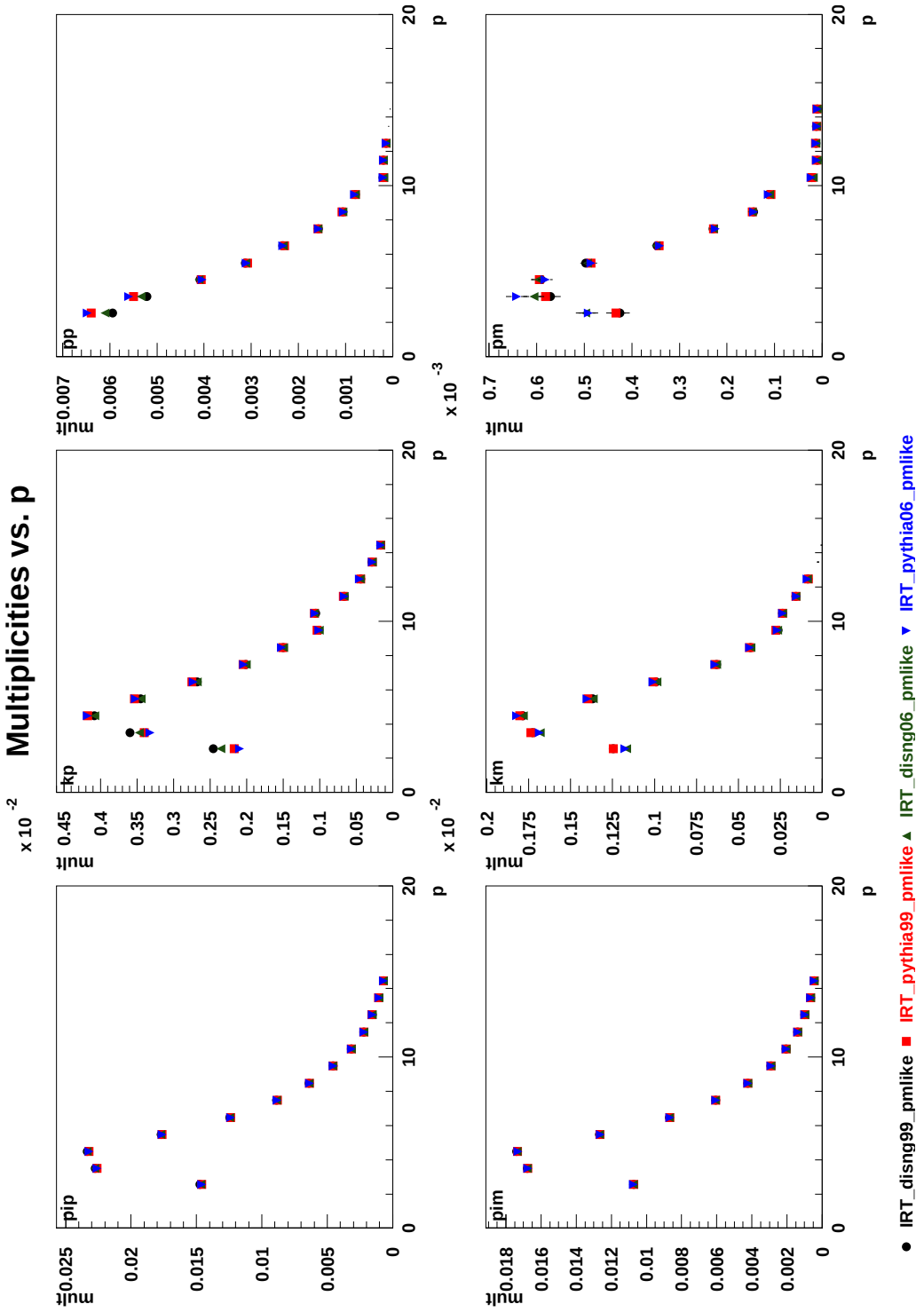

Figure 81: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using likeness separated IRT Pmatrices extracted from MC productions using different generators and geometries.

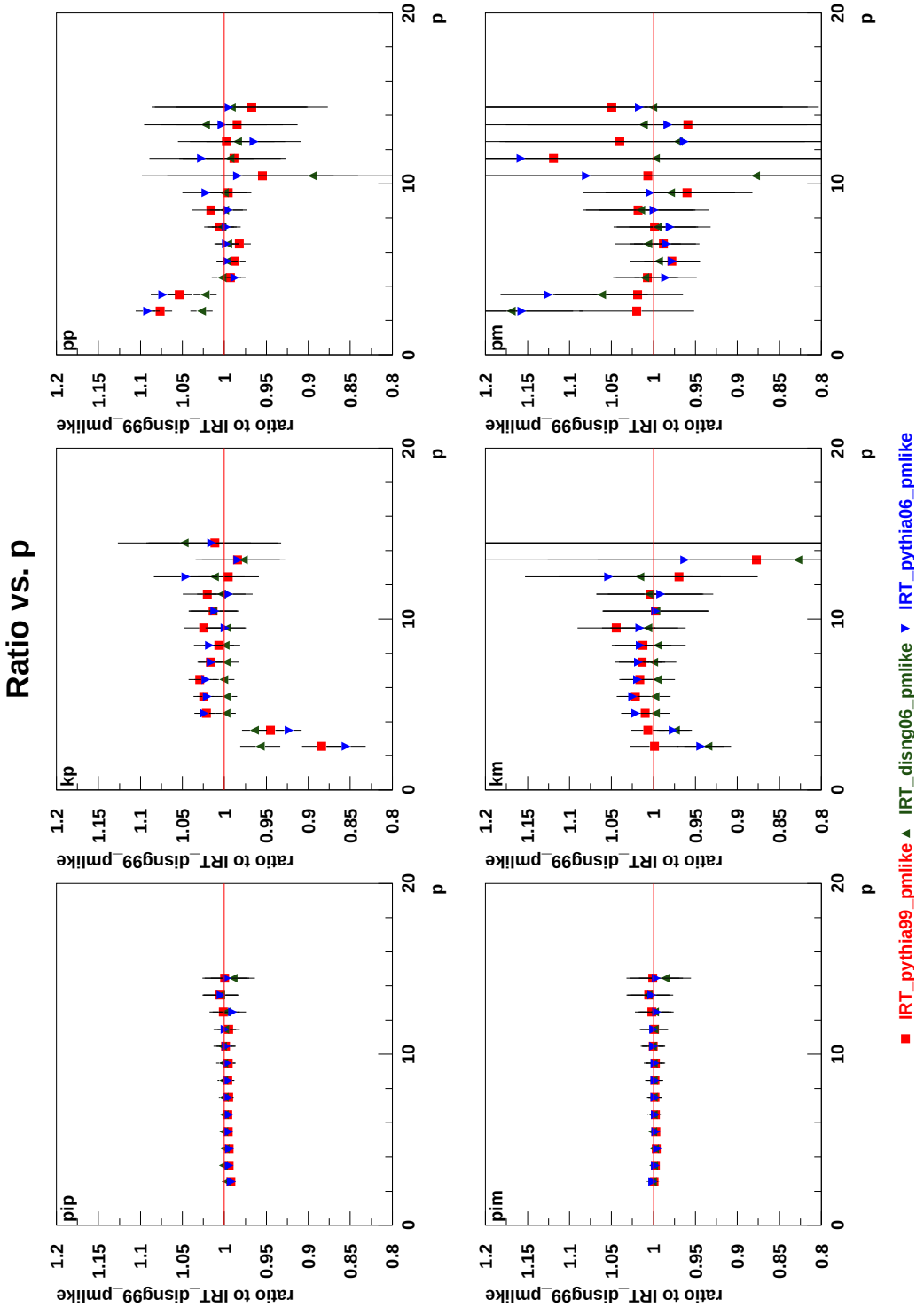

Figure 82: Ratios of the multiplicities shown in Figure 81, relative to the first data set.

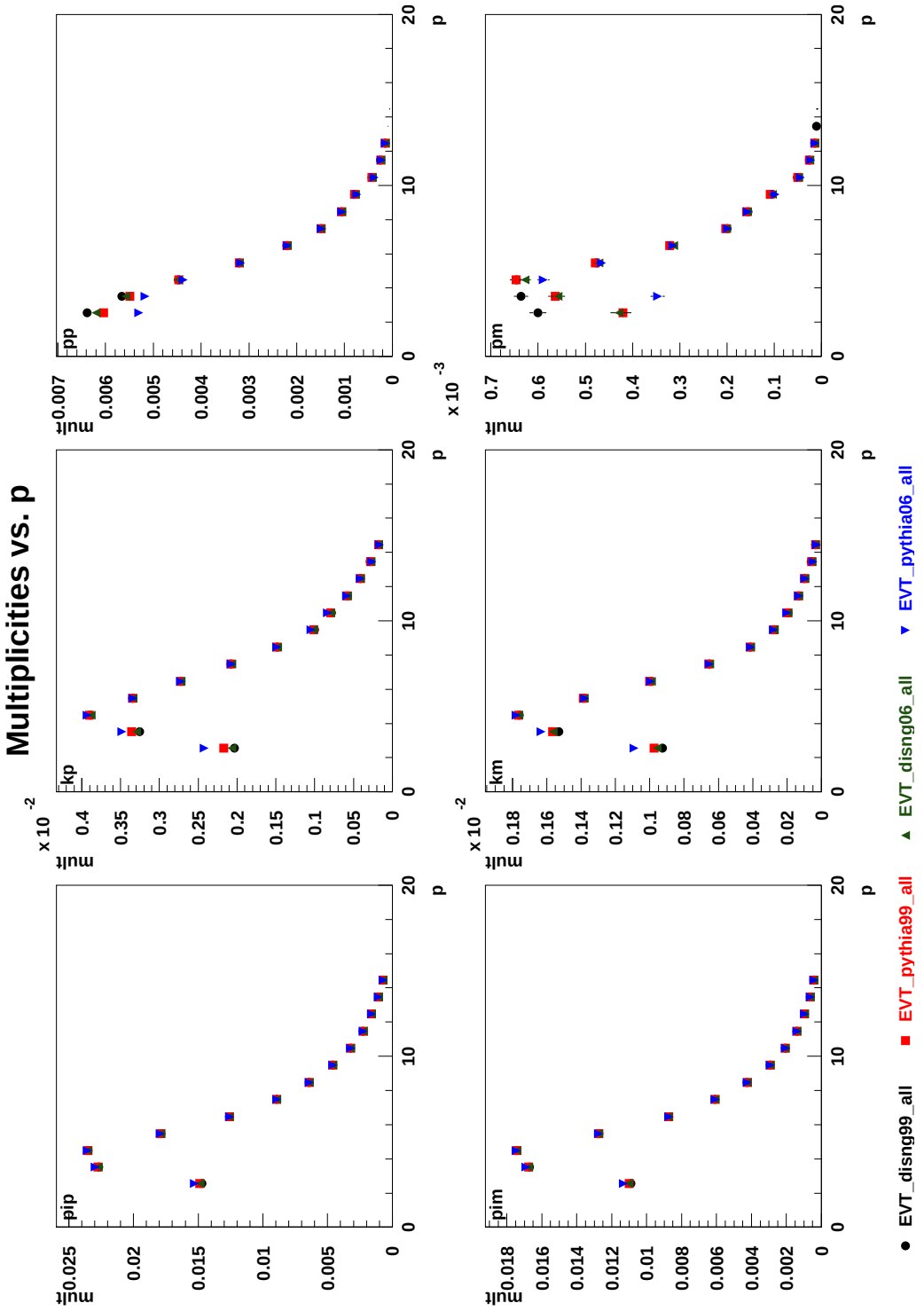

Figure 83: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using charge combined EVT Pmatrices extracted from MC productions using different generators and geometries, each of which were produced with their own background file.

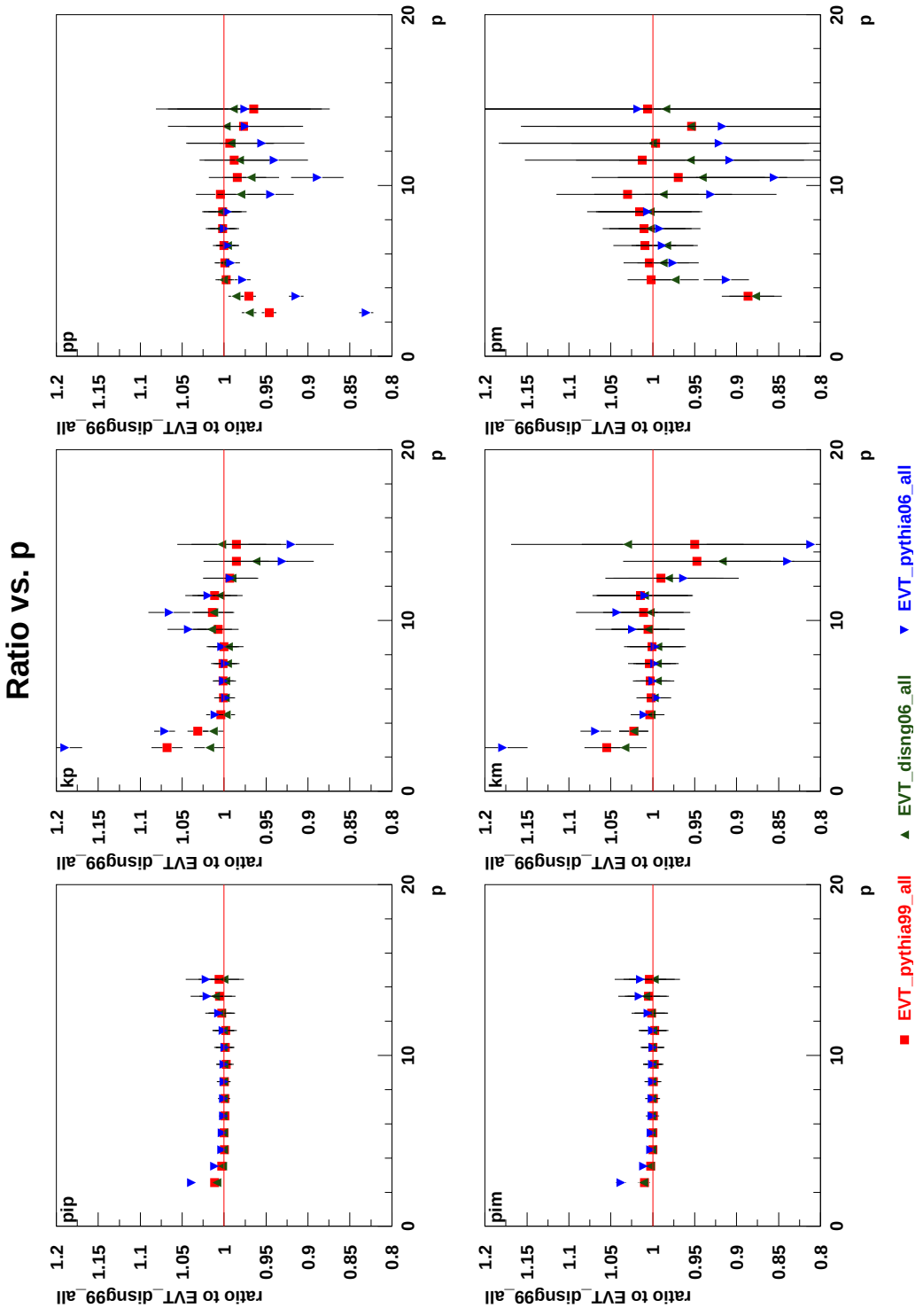

Figure 84: Ratios of the multiplicities shown in Figure 83, relative to the first data set.

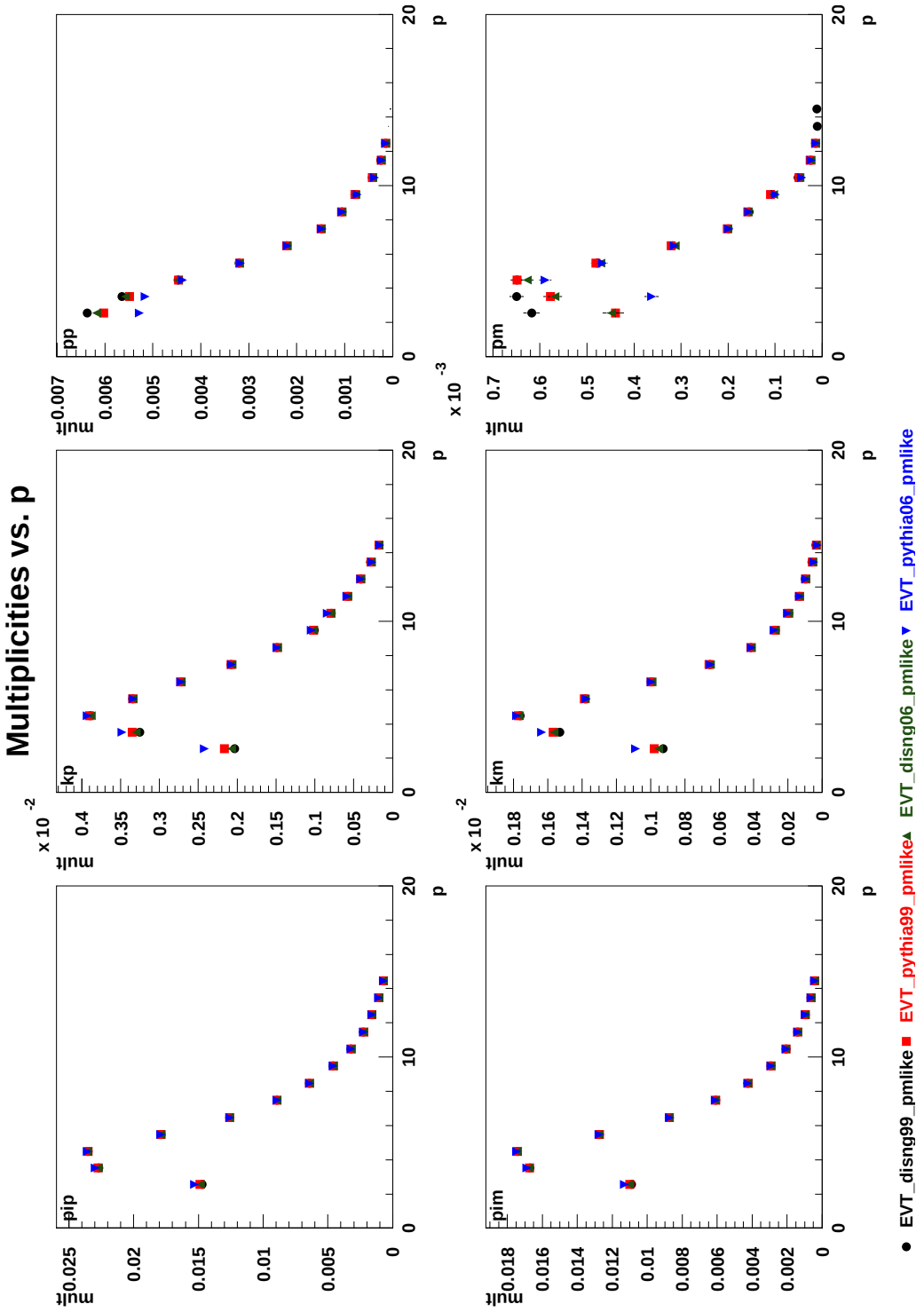

Figure 85: Multiplicities from unpol data from the 00d2 production. The multiplicities were unfolded using likeness separated EVT Pmatrices extracted from MC productions using different generators and geometries, each of which were produced with their own background file.

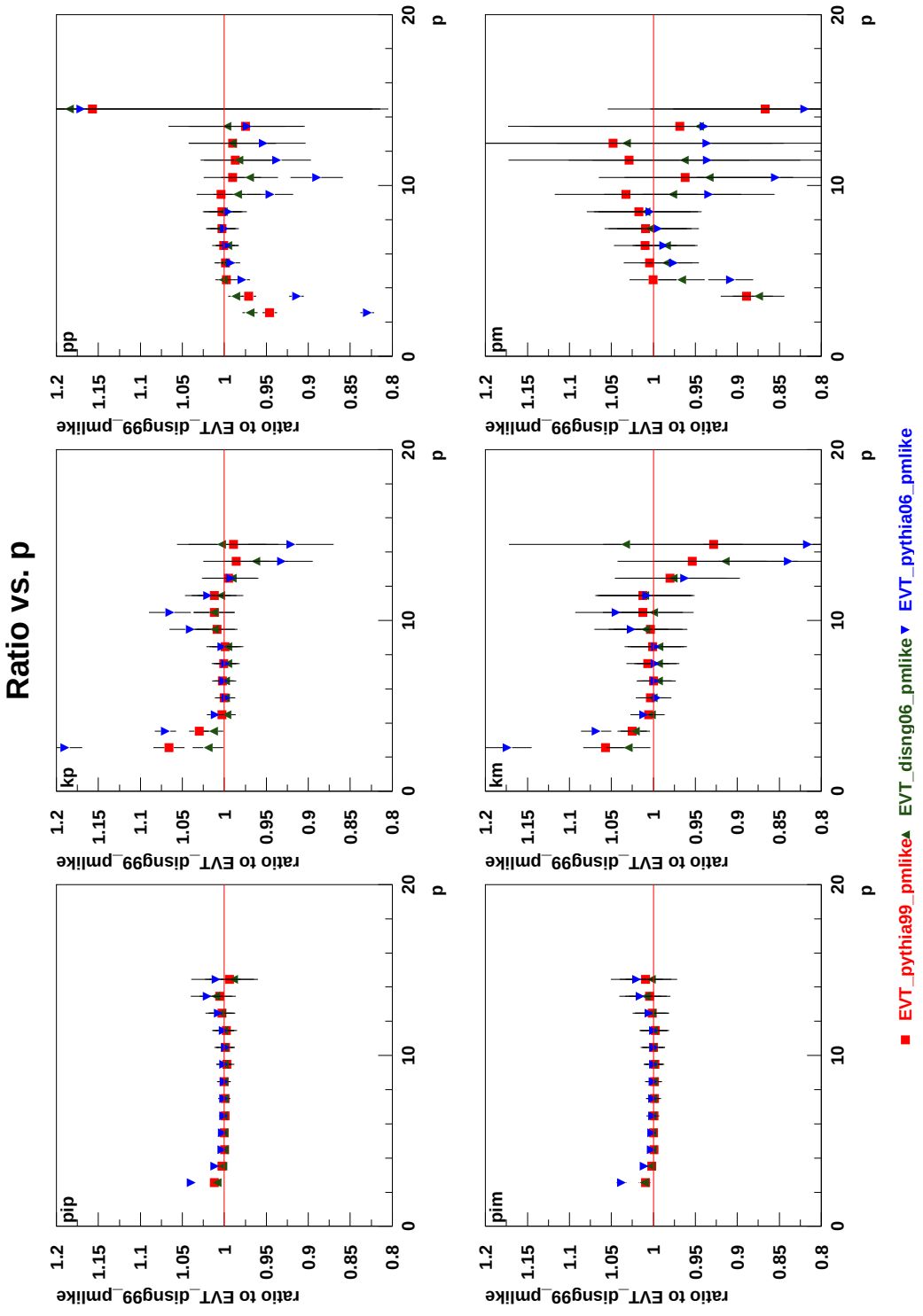

Figure 86: Ratios of the multiplicities shown in Figure 85, relative to the first data set.

List of Figures